VOLUME 102 PART B NUMBER 1 JANUARY 1955

74

[illegible] FEB [illegible]

The Proceedings

OF

THE INSTITUTION OF ELECTRICAL ENGINEERS

FOUNDED 1871: INCORPORATED BY ROYAL CHARTER 1921

PART B

RADIO AND ELECTRONIC ENGINEERING
(INCLUDING COMMUNICATION ENGINEERING)

GE/L7 As.

The Institution of Electrical Engineers

FOUNDED 1871
INCORPORATED BY ROYAL CHARTER 1921

PATRON: HER MAJESTY THE QUEEN

COUNCIL 1954-55

President

J. ECCLES, C.B.E., B.Sc.

Past-Presidents

SIR JAMES SWINBURNE, BART., F.R.S.
W. H. ECCLES, D.Sc., F.R.S.
THE RT. HON. THE EARL OF MOUNT EDGCUMBE, T.D.
J. M. DONALDSON, M.C.
PROF. E. W. MARCHANT, D.Sc.
P. V. HUNTER, C.B.E.
H. T. YOUNG.
SIR GEORGE LEE, O.B.E., M.C.
SIR ARTHUR P. M. FLEMING, C.B.E., D.Eng., LL.D.
J. R. BEARD, C.B.E., M.Sc.
SIR NOEL ASHBRIDGE, B.Sc.(Eng.).
COLONEL SIR A. STANLEY ANGWIN, K.B.E., D.S.O., M.C., T.D., D.Sc.(Eng.).
SIR HARRY RAILING, D.Eng.
P. DUNSHEATH, C.B.E., M.A., D.Sc.(Eng.).
SIR VINCENT Z. DE FERRANTI, M.C.
T. G. N. HALDANE, M.A.
PROF. E. B. MOULLIN, M.A., Sc.D.
SIR ARCHIBALD J. GILL, B.Sc.(Eng.).
SIR JOHN HACKING.
COLONEL B. H. LEESON, C.B.E., T.D.
H. BISHOP, C.B.E., B.Sc.(Eng.).

Vice-Presidents

T. E. GOLDUP, C.B.E.
S. E. GOODALL, M.Sc.(Eng.).
WILLIS JACKSON, D.Sc., D.Phil., F.R.S.
SIR GEORGE H. NELSON.
SIR W. GORDON RADLEY, C.B.E., Ph.D.(Eng.).

Honorary Treasurer

H. W. GRIMMITT.

Ordinary Members of Council

J. BENNETT.
A. R. COOPER.
A. T. CRAWFORD, B.Sc.
C. DANNATT, O.B.E., D.Sc.
B. DONKIN, B.A.
O. W. HUMPHREYS, B.Sc.
C. R. KING, C.B.E.
H. R. L. LAMONT, Ph.D., M.A., B.Sc.
F. J. LANE, O.B.E., M.Sc.
G. S. C. LUCAS, O.B.E.
G. LYON, M.Sc.(Eng.).
SIR HAMISH D. MACLAREN, K.B.E., C.B., D.F.C., LL.D., B.Sc.
A. H. MUMFORD, O.B.E., B.Sc.(Eng.).
W. F. PARKER.
PROF. M. G. SAY, Ph.D., M.Sc.
R. L. SMITH-ROSE, C.B.E., D.Sc., Ph.D.
G. O. WATSON.
J. H. WESTCOTT, B.Sc.(Eng.), Ph.D.
E. L. E. WHEATCROFT, M.A.
R. T. B. WYNN, C.B.E., M.A.

Chairmen and Past-Chairmen of Sections

Measurements:
M. WHITEHEAD.
*J. F. COALES, O.B.E., M.A.

Radio:
C. W. OATLEY, M.A., M.Sc.
*J. A. SMALE, C.B.E., A.F.C., B.Sc.

Supply:
J. D. PEATTIE, B.Sc.
*L. G. BRAZIER, Ph.D., B.Sc.

Utilization:
J. I. BERNARD, B.Sc.Tech.
*B. L. METCALF, B.Sc.(Eng.).

Chairmen and Past-Chairmen of Local Centres

East Midland Centre:
J. H. MITCEHLL, B.Sc., Ph.D.
*C. D. WILKINSON.

Mersey and North Wales Centre:
P. R. DUNN, B.Sc.
*T. COATES, M.Eng.

North Midland Centre:
W. A. CROCKER.
*G. CATON.

North-Eastern Centre:
G. W. B. MITCHELL, B.A.
*H. ESTHER, B.Eng.

Southern Centre:
E. A. LOGAN, M.Sc.
*COMDR.(L) C. V. ROBINSON, R.N., O.B.E.

North-Western Centre:
PROF. E. BRADSHAW, M.B.E., M.Sc.Tech., Ph.D.
*H. WEST.

Northern Ireland Centre:
MAJOR P. L. BARKER, B.Sc.
*J. R. W. MURLAND, B.Sc.(Eng.).

Western Centre:
A. N. IRENS.
*J. VAUGHAN HARRIES.

Scottish Centre:
J. S. HASTIE, B.Sc.(Eng.).
*C. H. A. COLLYNS.

South Midland Centre:
A. R. BLANDFORD.
*H. J. GIBSON, B.Sc.

* Past-Chairman.

RADIO SECTION COMMITTEE 1954-55

Chairman

C. W. OATLEY, M.A., M.Sc.

Vice-Chairmen

H. STANESBY.
R. C. G. WILLIAMS, Ph.D., B.Sc.(Eng.).

Past-Chairmen

J. A. SMALE, C.B.E., A.F.C., B.Sc.
E. C. S. MEGAW, M.B.E., D.Sc.

Ordinary Members of Committee

PROF. H. E. M. BARLOW, Ph.D., B.Sc.(Eng.).
N. R. BLIGH, B.Sc.(Eng.).
G. G. MACFARLANE, Dr.Ing., B.Sc.
B. N. MACLARTY, O.B.E.
J. H. H. MERRIMAN, M.Sc.
BRIG. E. J. H. MOPPETT.
H. PAGE, M.Sc.
A. R. A. RENDALL, Ph.D., B.Sc.
N. C. ROBERTSON, C.M.G., M.B.E.
L. RUSHFORTH, M.B.E., B.Sc.
J. A. SAXTON, D.Sc., Ph.D.
A. M. THORNTON, B.Sc.

And

The President (*ex officio*).
The Chairman of the Papers Committee.
A. H. MUMFORD, O.B.E., B.Sc.(Eng.) (representing the Council).
G. E. MIDDLETON, M.A. (representing the Cambridge Radio Group).
D. H. THOMAS, M.Sc.Tech., B.Sc.(Eng.) (representing the North-Eastern Radio and Measurements Group).
J. MOIR (representing the South Midland Radio Group).

The following nominees of Government Departments:
Admiralty: CAPTAIN(L) G. C. F. WHITAKER, R.N.
Air Ministry: AIR COMMODORE G. H. RANDLE, R.N.
Department of Scientific and Industrial Research: A. F. WILKINS, O.B.E., M.Sc.Tech.
Ministry of Supply: F. S. BARTON, C.B.E., M.A., B.Sc.
Post Office: CAPTAIN C. F. BOOTH, O.B.E.
War Office: COL. J. H. E. PANTON, O.B.E.

Secretary

W. K. BRASHER, C.B.E., M.A., M.I.E.E.

Deputy Secretary

F. JERVIS SMITH, M.I.E.E.

Assistant Secretary

F. C. HARRIS.

Editor-in-Chief

G. E. WILLIAMS, B.Sc.(Eng.), M.I.E.E.

EᵗᶜD 15 1958-59

No 10 [illegible]

COLLEGE OF ENGINEERING
1 DEC 1958

The Institution is not, as a body, responsible for the opinions expressed by individual authors or speakers

THE PROCEEDINGS OF
THE INSTITUTION OF ELECTRICAL ENGINEERS

EDITED UNDER THE SUPERINTENDENCE OF W. K. BRASHER, C.B.E., M.A., M.I.E.E., SECRETARY

VOL. 102. PART B. No. 1. JANUARY 1955

Paper No. 1804
Oct. 1954

INAUGURAL ADDRESS

By J. ECCLES, C.B.E., B.Sc., M.I.C.E., M.I.Mech.E., President.

(*Address delivered before* THE INSTITUTION *7th October*, 1954.)

A sense of inadequacy serves to heighten my feeling of gratitude to my fellow-members for having accounted me first amongst equals in The Institution for the period of one year. My limitations are obvious and well known to many. As I stand here somewhat in awe of those giants, my predecessors, some of whom adorn these walls and benches, I can but promise to do my utmost to serve, whilst you may derive comfort from the thought that the wisdom of the Council, the knowledge of the Secretariat, and the restraining hand of tradition, will protect your interests. I thank you all.

The Institution has a membership of nearly 39 000, but mere number is not strength, and the strength of this Institution derives from the fact that it knits together in common fellowship those professionally engaged in the whole range of electrical engineering, from the practical application of well-established technology to the translation of the newest theory into the technical marvels of our age.

To do so, there must be wise and energetic leadership with the maximum practicable degree of devolution of responsibility to the Sections, Centres and Sub-Centres. That this is achieved is a tribute to the enlightened guidance of Mr. Brasher and his staff, and to the reservoir of managerial experience distilled through the deliberations of your Council. The manner in which the Sections and Centres discharge their duties, and seek to promote the highest interests of The Institution and the technical well-being of its members, is a source of constant admiration.

Through the papers that are published, with or without discussion, The Institution seeks to include in its *Proceedings* a complete record of technical progress in all branches of electrical engineering. Many of these papers are necessarily of a highly specialized and mathematical nature. Whilst recognizing that a high degree of specialization is essential to economic success and that the tendency may increase, it is important to ensure that the general practitioner and other specialists are kept informed of the trend of events in fields other than their own. To this end, it is greatly to be hoped that more members will find time to present papers of a general character on the current techniques in their own spheres.

As I hope to show, we are faced with the prospect of an ever-expanding range of research and experiment in difficult fields, and if we are to succeed it is imperative that the volume and quality of technical education and training should be stepped up. It is the privilege and the duty of The Institution to offer informed advice on this matter to those whose responsibility it is to decide the form and extent of the educational facilities to be provided. This duty is being discharged.

One of the things causing concern is the inadequacy of the intake of students and juniors to the electrical engineering profession. Insufficient schoolboys are imbued with scientific insight and knowledge of the part that science must inevitably play in securing an acceptable existence in an industrial country. Too many are unable to pass the university entrance examinations. If this defect should stem from lack of appreciation by headmasters, housemasters and careers masters of the indispensable necessity of scientific education, it calls for immediate remedial action at those levels. If the teaching of science in schools cannot be improved and extended, the universities may have to consider the selection of entrants more on potentialities than on detailed scientific knowledge, and themselves teach science from an earlier stage. Character and an inquiring mind may be a better foundation than very early knowledge of Ohm's law.

A further thought on this may be worthy of record. It is a fact that each generation of technologists has to learn its technology from the beginning. There is no way of starting where the other man left off. Each must plod through the A B C of the subjects he elects to study. To simplify this process should be a duty of educationalists, and how better might this be started than by simplifying the units in which basic quantities are measured? The rationalized M.K.S. system of units has been recommended as an international standard by leading authorities, and it would seem that the time has now come when this simplification could be used exclusively in textbooks and academic courses.

Although there is no restriction upon the subject-matter of an Inaugural Address, it will not have escaped notice that almost invariably each holder of this office has, quite naturally, chosen to discuss in his Address some aspect of the work upon which he is engaged.

I am an electrical power engineer and, fortunately for you, though unhappily for me at this moment, Sir John Hacking in his memorable Address three years ago, and more recently in his Presidential Addresses to the British Electrical Power Convention and to the Junior Institution of Engineers, has given very full accounts of the immediate past and middle-distance future of electrical power engineering in this country.

One could attempt to dot some of the i's and cross the t's of what has already been written. The story of the 275-kV Grid could be brought up to date. Then, too, there is the proposed cross-Channel connection linking Britain with the European mainland across a 26-mile stretch of turbulent sea with rocky

---

Mr. Eccles is a member of the British Electricity Authority.

foreshores and an uneven bottom reaching a depth of 216ft at its lowest point. Again, there is the continued story of endeavour and achievement in the field of power generation in which the equivalent of six modern power stations are fully equipped each year, thus bringing into use some one and a half million additional kilowatts of electrical generating capacity in Great Britain alone. On the utilization side there is a human story in the manner in which electricity is spearheading the drive for food production and arresting the depopulation of the countryside, and there is much of technical interest in the means whereby annually over 10 000 farms and three times as many rural dwellings are receiving supply in this country for the first time.

It would have been possible to develop and arrange thoughts on these and other aspects of electrical power engineering, but inevitably they would have had to be of a detailed character and so were better dealt with in discussable papers.

In this Address, therefore, I wish, with your permission, to depart a little from tradition and offer some observations of a very general character on the long-term past and future of power engineering. Much of what I propose to say will be familiar to many members, for never before has there been such an awareness of the consequences of continued exploitation of known natural sources of energy or so much written about the need to discover and utilize new ones. From the initiated I crave indulgence.

Before delving into the past or peering into the future, it may be well to consider what it is the power engineer seeks to do and why it is necessary to have him around at all. The function of the power engineer, as I understand it, is to give man command over Nature by releasing her great storehouse of energy and presenting it in forms and quantities that, at one end of the scale, enable him to achieve the otherwise impossible, and at the other, enable him to gain a livelihood without undue physical exertion. Between these extremes he provides, in a great variety of ways, the means whereby man may live a fuller life, or if he chooses, effect his own destruction. To-day we accept the benefits as a matter of course, and realize their significance only when they fail to appear, but let us pause for a moment to discuss their magnitude and their effect upon our mode of living.

Each year in this country we consume about 200 million tons of coal. It is probable that half of this is used to produce mechanical and electrical power. Taking a realizable efficiency of conversion from coal to usable power at the point of application, it is easy to show that 100 million tons of coal used in this way will do the work of at least 750 million unaided men working a 44-hour week for 50 weeks. The population of this island is rather less than 50 million, and hence by the conversion of this amount of coal into work we have endowed each man, woman and child with fifteen slaves to do their bidding. These slaves propel trains and trams, operate cranes, turn lathes and boring mills, whirl vacuum cleaners and washing-machines and actuate the myriad host of power-driven appliances on which our civilization depends.

Oil provides another set of slaves to propel cars, buses and lorries, tractors and bull-dozers, ships and aircraft, and a calculation of the same kind shows that the present consumption of liquid fuel in this country provides the equivalent of a further five slaves for each living soul in Great Britain. These latter have to be purchased in, and transported from, foreign lands, but, this done, the whole platoon requires neither food nor raiment, and all we have to do is to provide their working tools and ensure that they are usefully employed.

The conversion factors used may not be quite accurate, but they are of the right order, and the result illustrates our utter dependence, in the 20th century A.D., on the sources of power in nature brought under control and regulated for our benefit by the power engineer.

How then has it come about that each of us has to depend upon 20 inanimate slaves to maintain his present standard of mobility, comfort and leisure? For how long has this been going on, and what are the chances of its continuance? These are some of the vital questions of our age, and in attempting to answer them let us go backwards in history a little way.

Some geologists compute the age of the earth to be about 3 000 million years. Nine-tenths of that period passed without any event of significance to power engineers until about 300 million years ago, in circumstances which we can only conjecture, there seems to have occurred upon this planet a condition, or series of conditions, conducive to luscious vegetable growth over large tracts of its swampy surface. Vegetation grew, decayed, was reborn, and decayed, repeating its life cycle for perhaps tens of millions of years, until there was built up stratum upon stratum of hydro-carbonaceous matter.

Times changed, the climate altered, there were violent internal explosions with consequent disturbance of the earth's crust, dry land became the sea-bed, the valleys were exalted, the mountains were brought low, there were long periods of snow and ice, and never again have there been repeated in nature conditions so conducive to an accumulation of her precious hydrocarbon store. To this event, or series of events, we owe the coal measures and the oil wells which to-day provide over 85% of the world's fuel. This garnering of nature's harvest being over, there is nothing of significance to relate until beings resembling man made their appearance about a million years ago. We know little that is not somewhat speculative about the evolution of man, before the period of recorded history. It is thought, however, that a million years ago he knew how to initiate and control fire, and it would seem that his progress since then has been conditioned as much by the state of his technology as by any other single cause. For a very long time men remained gatherers as distinct from producers. They were content to take what nature provided and use it as best they could. Later they assumed a more active role and endeavoured to improve upon nature by tilling and fertilizing the soil, using domesticated animals for work and food and fashioning tools of stone and metal to increase the range and usefulness of their endeavours. It was in this latter phase that the demand for fuel grew apace. Wood was required for warmth, for cooking and for smelting metal, and whole districts were deforested by burning for no other reason than to provide wood-ash as a fertilizer. Large areas in India, China and other countries were devastated in this way, and when in these areas crops failed through lack of further enrichment of the soil, the population was decimated by starvation, and so the culture died that had blossomed in the previous age of plenty.

The record of these early civilizations seems to show that an improvement in technology encouraged an increase in population which overshot the capacity of the reigning technology to feed and to clothe, with the result that, in the exercise of that strongest human instinct—self-preservation—man with his limited knowledge unwittingly cut off the branch on which he sat. Thus it was that lack of knowledge of what nature held in store, and lack of knowledge of how to use efficiently the known resources, caused cultures of high rank to wither and, in some cases, to vanish from the earth.

However, the pattern was not even and, in some areas, improving supply technology, aided by slave labour, was able to meet the increasing demand for long periods. By 10000 B.C. men knew how to make bricks, by 6000 B.C. they began to tap the fossil fuels, beginning probably with the most accessible deposit, asphalt. Earthenware vessels were made as early as 5000 B.C. The Chaldeans were skilful metal workers by 4000 B.C.

Limestone was calcined and enamelled pottery made in Egypt and Babylon by 3000 B.C. Asphalt was used in the building of the Pyramids and for waterproofing the walls of Jericho. In Babylon, roads were built by setting stones in asphalt, a practice revived by J. L. McAdam some 2 500 years later.

About 1000 B.C. the Chinese were using natural gas for fuel and lighting, which they procured from wells 3 000ft deep and transmitted in bamboo pipes. According to Herodotus, an oil well was working on the island of Zante in 400 B.C., the lighter fractions being separated by stretching a hide over a cauldron of boiling oil and wringing out the condensed liquid.

The earliest recorded use of coal is 1100 B.C. in China. At that time, Chinese technology was sufficiently advanced to enable them to produce saltpetre, arsenic, mercury, vegetable oils, paper, sugar, printing and gunpowder. The first use of coal outside China was probably in Greece, and there is evidence that coal was known and used in Britain before the Roman occupation.

These apparently isolated but perhaps not unconnected examples of technological progress coincided, in some cases, with quite advanced civilizations, but it may be significant that almost invariably an improvement in supply technology was accompanied by an increase in population and by an arrogance or indolence in the governing body, which formed the seed-bed of decay. In this respect, things came to a head round about 500 B.C. In 538 B.C., Belshazzar was slain on the night of the handwriting on the wall and the Babylonian empire came to an end. In China, industrial progress ceased with the full flowering of the Confucian philosophy, and in India the caste system withdrew all intellectuals from the industrial crafts. Greek industrial progress died with Alexander the Great.

These events and others ushered in a period of two thousand years of technological stagnation in which, whilst fuel was still used for heating, little further progress was made in establishing man's command over nature. This period includes the Greek and Roman civilizations, which made little attempt to revive fuel technology. Greek intellectuals scorned those who tried to turn technical theory to practical advantage, and, though the Romans invented the water-wheel and the concave mirror for concentrating solar heat, their source of energy was almost wholly that of slave labour and animals right to the end. What was needed was the discovery and application of means to augment man's energy and improve his transport. The Phoenicians had used the force of wind in sailing ships since 1000 B.C.; Hero of Alexandria came near to converting heat into work; the Romans used water power and the Chinese had used the explosive force of gunpowder. But no one knew how to harness nature continuously on the scale necessary to suit the growing human need.

It would be wrong to suggest that nothing happened anywhere during the Dark Ages. At the end of the Roman occupation, Britain experienced a political setback, but with the gradual emergence of more settled times the demand for fuel outstripped the current rate of vegetable growth and the countryside was being steadily denuded of its forests. Long before this, coal had been discovered as an outcrop and its value as a fuel was known. It was not a popular fuel, and in the reign of Edward I a decree was made compelling "all but smiths to eschew the obnoxious material and return to the fuel they used of old."

Thus we arrive at the 15th, 16th and 17th centuries A.D., with science and technology slowly awakening and facing the prejudices that had become entrenched in men's minds, without opposition, during their long period of hibernation. It was in this setting at the very end of the 17th century that an Englishman, Thomas Savery, invented the first successful steam engine and showed how heat could be converted into work. By means of this engine, improved by Newcomen and Watt, there was tapped an enormous reservoir of energy stored in the coal measures for 300 million years. The amount of power that was there for the taking seemed unlimited and the scale of human achievement made possible by this new mechanical aid dazzled the imagination. It was natural that the first applications of steam engines should be for pumping and winding at coal mines, but, once this main source of power energy had been made secure, their use became general. In 1827, the first steamship crossed the Atlantic to the New World, and in 1829 the first steam-driven railway locomotive was in operation. In 1884, Charles Parsons invented the first practical steam turbine, which so raised the sights in size and efficiency that to-day single machines are being designed for an output of 200 000 kW using only three-quarters of a pound of coal per kilowatt-hour.

Contemporary with these events in the realm of steam, work was in progress on the design of machines in which controlled explosions of oil or gas could be translated into continuous mechanical power. The result was the internal-combustion engine, which, apart from many useful fixed applications, completely revolutionized the mode, range and speed of transport. The first motor-car using this engine took the road in 1885, and the first aeroplane left the ground in 1903.

Prior to the emergence of most of these applications, Michael Faraday had discovered and demonstrated how a new and more versatile form of power could be made available. As every member of this Institution is aware, it was at the Royal Institution, London, on the 29th August, 1831, that Faraday first demonstrated the basic principle of electromagnetic induction now used in every electric power generator throughout the world.

The result of these epoch-making inventions and discoveries has been that man, after surviving for nearly a million years on the gifts of nature as they were currently produced, and being frustrated from further development by the inadequacy of his technology, has in some parts of the world during the past 250 years developed a method of living that transcends everything hitherto achieved, but is largely dependent upon the use in enormous quantities of natural energy stored in a bygone age.

Earlier in this Address, I suggested that the present use of energy was equivalent to the service of twenty slaves for each person on this island, and I posed questions on the chances of this state of affairs being maintained and what is likely to happen if the supply fails. I shall now attempt a qualitative answer to these questions in the light of current knowledge.

Coal, oil and natural gas are wasting sources of energy, and in due course will cease to exist. It has been estimated that, in this island already, we have raised 25 000 million tons of coal and that the readily accessible coal remaining is about twice that amount. On this basis and at a modest increase in consumption we shall exhaust the accessible coal in about 200 years. The total coal reserve is probably very much greater, but much of it would be extremely costly to win.

It is not easy to estimate the total coal reserves of the world, and although attempts to do so have been made, the results are subject to many qualifications—the thickness of the seam, the quality of the coal, its depth below the surface and many other things have to be evaluated before its economic worth can be assessed. Economic worth is a relative term which depends on the availability and cost of alternative fuels. Liquid fuel enjoys a somewhat special relationship to all others.

A recent American assessment of the world reserves of all fuels is shown in Table 1. This estimate suggests that the world reserves of crude oil are only about 5% of world coal reserves, whilst, with 1 : 1 breeder reactors, the potential energy of nuclear fuel is more than twenty times as great as that of the world reserves of coal, oil and natural gas put together.

Table 1*

Fuel	World reserves	Source of data	Total energy B.Th.U.
Crude oil	610 × 10^9 barrels	Weeks and Moulten	× 10^{18} 3·5
Natural gasoline	11·5 × 10^9 barrels	American Petroleum Institute	0·07
Shale oil	620 × 10^9 barrels	Bureau of Mines	4
Natural gas	560 × 10^{12}ft^3	American Gas Association	0·6
Coal	3 482 × 10^9 tons	Bureau of Mines	72·2
Total conventional fuel			80
Uranium ..	25 × 10^6 tons	Raw Materials Division of A.E.C.	1 700 at 1 : 1 breeding
Thorium ..	1 × 10^6 tons	do.	71
Total new fuel (say)			1 800

* Reproduced from CISLER, W. L.: "Economic Evaluation of the Industrial Use of Atomic Energy" (American Power Conference, March 1953).

The world annual consumption of coal is about 1 600 million tons and the consumption of crude oil is about 640 million tons. Unfortunately, simple arithmetic is of little practical value in computing the probable life of respective fuels in particular countries. There are some undeveloped countries with considerable coal resources in which the present use is negligible. On the other hand, highly developed countries are consuming coal at a rate disproportionate to their reserves. Furthermore, not all the known reserves are recoverable, although, on the other hand, more may be discovered.

With regard to oil, it has been estimated that consumption in the United States will be doubled between 1950 and 1975, and if this rate of increase were general the life of world reserves would be very much reduced. In view of the widespread tendency to use oil more extensively in agriculture and for sea, air and road transport, the supply of this fuel may easily become critical within the lifetime of the present rising generation.

This all too brief survey lends colour to the thought that, of the fossil fuels, oil will be the first to go. The efforts to produce a substitute liquid fuel may include the synthesis of coal, which, in turn, would accelerate the rate of destruction of this fuel. Even so, the reserves of coal are sufficient to allow time for the orderly development of alternatives. The geographical disappearance of fossil fuels will depend on the rate of uplift in relation to the amount of local reserves, and the rate of development of alternative fuels will depend, in the first instance, on their cost in comparison with that of local or imported fossil fuel. Obviously, the country with large reserves of all kinds of fuel is most favourably situated to maintain and expand its mechanized civilization, and provided it uses these resources economically and makes commensurate advances in its political and cultural life, it is probably destined to remain a world force for a longer period than other less well-endowed peoples. The importance, therefore, of the economical use of native fuel stretches far beyond the desirability of maintaining one's standard of living, although it includes it.

What then are the alternatives to fossil fuels? The general answer is nuclear fuel. I think, however, it would be wrong to give the impression that, although nuclear energy is there for the taking, it is a simple matter to release it in quantities suited to our requirements. Hitherto heat has been obtained by amputating and re-grafting the limbs of matter; in nuclear fission, it is proposed to tear out and divide its heart.

It is not possible to make an accurate forecast of what the future fuel situation will be. There are so many imponderables, the frontier of knowledge is being constantly expanded, Nature herself may take a hand as in the previous Ice Ages, and the reaction of humans to changing circumstances must be given due weight.

The increase in world population is an important factor. Ever since man's emergence from the Dark Ages, world population has been increasing. Curiously enough, this has been noticeable even in countries where technology showed no advance. Perhaps the general trend has been made possible through better means of communication whereby the effects of industrial technology and medical science are felt beyond the confines of the countries that initiate them. Whatever the reason, most countries have participated in the production of a world population which has enjoyed a four-fold increase during the last 300 years and is now growing by 20 millions a year—that is by more than 2000 an hour.

The fate of previous civilizations has been sealed when population outstripped the ability of their technology to provide sustenance and shelter, and the fate of our own cannot be dissociated from this self-same influence.

In this welter of possibilities, it may be of value to look at two possible alternative situations—one on the assumption that nuclear fission or nuclear fusion will not prove to be a practical means of providing power, and the other that it will.

Without atomic energy, the main alternatives to coal, oil and natural gas are water power, tidal power and wind power, with solar energy and geothermic heat available in certain zones, and wood, peat, and animal wastes each making its own contribution.

Water Power

The conditions which make water power most readily available are adequate rainfall on a large elevated catchment area, a suitable reservoir in which to impound water at this high level, a relatively steep conduit to a low level where the power unit is situated, and adequate means for disposing of the spent water.

Hitherto the assessment of world inland water-power resources has included only those which show prospect of economic development and are not required for other purposes such as navigation, irrigation or domestic use. In future, much more may be pressed into service, perhaps for dual purposes of which power will be one. However, the most recent estimates show that the world total of inland water power, 40% of which is in Africa, could supply at least three-quarters of man's present energy requirements.

One of the largest and most spectacular sources of water power lies at the western end of the Mediterranean. This inland sea loses water through evaporation which is not wholly recovered from its river inflow, with the result that water is constantly flowing through the Straits of Gibraltar, and to a lesser extent through the Dardanelles. If, therefore, both straits were dammed, the level of the Mediterranean would gradually fall, and having established a suitable differential it could be maintained by admitting water from the Atlantic through suitable turbines. Rough calculation shows that, with a differential of 70ft, the inflow could develop 12 million kW, which is 75% of the present demand for electricity in Great Britain. It would require a modern Hercules to undo the work of his illustrious namesake, and no doubt the political repercussions would be considerable. The project is mentioned here only to illustrate the kind of thing that may have to be done in a more civilized world of the future when coal has ceased to be.

Tidal Power

Tides are produced by the combined gravitational effect on the earth of the moon and the sun. Since the liquid portion of the earth's surface is free to move, large oceans respond to the differential attractive force at the nether and the further shore. Winds and currents also play a significant part. The result is that, at any given point on the coastline of a large ocean, the tidal lift normally varies cyclically twice in the course of a lunar day, twice in the lunar month and twice a year. The amplitude of tidal lift varies from place to place. Since tides follow the moon and man's efforts follow the sun, the tidal cycle is not co-ordinated to human needs, and, in this sense, tides are an unsatisfactory source of energy.

The simplest device for using tidal energy is to trap water in an estuary, or reservoir, at the time of high tide, and after the tide has receded sufficiently allow the trapped water to escape through turbines in which most of the energy due to the differential head is converted into useful work. The daily work periods are cyclic, and their incidence in time of day also traverses twelve hours during each lunar month. There are devices, such as pumped storage and double-basin working, by means of which the energy available for external work can be controlled and better fitted to human requirements. There is also a difference of up to two hours in the time of high tide at different places round our coast. If a powerful electrical transmission system were available to connect together all such points, it would be possible to take advantage of these time differences to reduce the amount of pumped storage. So long as the harnessed tidal energy formed a small proportion of the total energy in the electricity system to which it was connected, it could be accepted as and when it was available without storage. The largest single tidal scheme in this country is that known as the Severn Barrage, and much thought has been devoted to the most effective way of harnessing the energy in this estuary. A number of comparable schemes are being investigated in other countries, but no tidal-energy project has yet reached the construction stage.

Energy of the Wind

Force is required to move air from one place to another, the corresponding energy being proportional to the cube of the velocity. It is practicable to abstract some of this energy by slowing down the speed of air movement.

Because of the low density of the medium, the energy release per unit volume for speed changes likely to be achieved in any practicable windmill is quite small. Stated conversely, a windmill must be a bulky structure with large-diameter vanes for even the relatively small output of 100 kW contemplated in recent designs. Furthermore, the wind bloweth where [and when] it listeth. It would seem that, although there is a large amount of energy of movement in atmospheric air treated globally, practical considerations will limit the contribution to be made by this source to a small proportion of man's present-day energy requirements. However, even this contribution may be welcome when coal and oil have disappeared.

Solar Energy

The amount of solar energy reaching the land areas of the earth's surface is equivalent to 10 000 times man's present requirements, and this may be the source from which the final scientific civilization will obtain its energy. Unfortunately, owing to the moderate temperature, the low intensity and the daily and seasonal variations, there is little hope of producing mechanical power by present known methods in any but tropical and semi-tropical regions. Interesting developments are taking place in domestic space-heating and in solar cooking. Some 6 000 solar cookers are being sold in India per annum, each capable of cooking a vegetable meal in about twenty minutes. If this practice extends, hundreds of millions of people will be able to cook their daily meal all the year round without using animal wastes for fuel. Soil fertility should improve correspondingly.

Neglecting, therefore, for the moment the possibilities of nuclear fission, the most reliable and economical alternative to coal and oil is water power if one looks at the world as a whole. However, the locations of possible water-power projects are such that quite fantastic power transmission schemes would be required to transport the energy to present centres of population. Apart from the political implications of such energy transfers across national frontiers, they present technical problems of a scale and type quite beyond anything that has been solved hitherto.

An alternative to transmitting the energy to the people is to transport the people to the energy source, and a considerable shift in the weights of population may well take place in the future if nuclear fission proves to be intractable as a source of industrial power. Great efforts would, no doubt, be made to ease the situation by developing tidal, wind and solar power in localities where there was little water power.

The foregoing assessment of the situation which may arise in the absence of a successful outcome of the attempts to harness the fission or fusion process for industrial purposes suggests that, when coal and oil are exhausted, it should be possible to muster man's present-day energy requirements from all the known sources, but that energy transmission or population transportation would present enormous technical and social problems. It underlines the absolute necessity to master the fission or fusion technique if the present pattern of civilization is to endure.

Nuclear Energy

Let us now look briefly at the nuclear-power possibilities.

Professor Einstein postulated that matter could be converted into energy. He calculated that the annihilation of one pound of matter would release energy equal to 11 340 million kWh of electricity. Were such complete conversion possible, six pounds of matter would release energy equivalent to the whole of the electricity generated in Britain last year. No means has yet been found to achieve this result.

It has been possible, however, to release a tiny fraction of the mass energy of matter by persuading a heavy atom to divide into two lighter atoms whose combined mass is a little less than that of the heavy one. Similarly, it has been possible to release energy by persuading a number of light atoms to form a single atom whose weight is a little less than the sum of the weights of the lighter atoms which combine to make it.

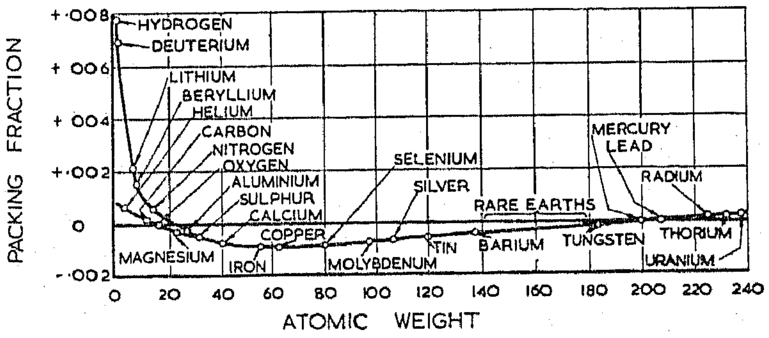


Fig. 1

Reproduced from KENDALL, J. T.: "The Production of Atomic Energy," *Engineering*, 1946, 161, p. 75.

Fig. 1 illustrates some of the possibilities. In it are shown the deviations from whole numbers of the atomic masses of a number of elements on a scale in which the mass of oxygen is taken as 16. Some of these deviations are positive and some negative. Theoretically there should be a release of energy when atoms with a more positive mass deviation divide or combine to produce atoms of a less positive or of a negative mass deviation. The division of uranium into two lighter elements or the fusion of hydrogen into helium offers the possibility of energy release. The diagram suggests that there are other possibilities.

Since hydrogen constitutes a substantial proportion of the earth's crust, the potential energy from the hydrogen-helium fusion process could be enormous. A practical difficulty is the sustained temperature at which this reaction takes place, and it would seem that, until something is discovered which will enable the reaction to proceed at a lower sustained temperature, the industrial application of this great energy source will remain unsolved.

From published data on the fission of uranium, it would seem that the process can be controlled in such a way as to liberate heat at a temperature suitable for industrial power purposes, but that before efficient large-scale nuclear power stations are practicable, work has to be done on heat-transfer methods and materials and on the treatment of the products other than heat which form part of the process. There can be little doubt that these problems will be solved if the need to do so is sufficiently great.

Of the uranium existing in nature, only about 0·7% is fissile. However, a process has been devised through which the remaining 99·3% can, by a double reaction, be converted into plutonium, which is fissile. Thus it is now theoretically possible to release in a controlled manner the energy of fission of all the uranium and perhaps the thorium content of the earth's crust. The practical and economic problem is to find the uranium and to separate and purify it.

From this point of view, it is unfortunate that uranium is so widely dispersed amongst the other materials of the crust. For example, it has been stated that the uranium content of a ton of granite, if it could be isolated, would produce as much heat as a ton of coal. However, there is no known economical method of isolating the uranium content of granite, and this is equally true for trace contents in many other common substances.

There remains, therefore, the problem of recovering pure uranium and thorium from their ores where they appear in forms and quantities that render the process economically practicable. It has been estimated that the availability and accessibility of uranium is such that (at present costs) an expenditure of £50 per pound of pure metal should be sufficient to mine and refine enough of this fuel to provide man's present energy requirements for 1 500 years. Similarly, on the same basis, an expenditure on less productive ores of up to £100 per pound of pure metal should procure sufficient uranium to last another 8 500 years. These very general figures should be taken only to indicate the satisfactory manner in which uranium can meet world energy requirements and, broadly, what effort would be required to render it available.

It is outside the scope of this Address to forecast the international problems that may arise from the geographical distribution of fissile material in nature, but without wishing to minimize them it can be said that they will just have to be solved if all peoples are to enjoy the benefits of a mechanized civilization.

There are some who argue that, because it is now theoretically possible for nuclear fission to satisfy man's total demand for energy, the pursuit of all other sources should be abandoned. I do not share that view. There are problems to be solved before the energy of fission becomes a reality on a large scale. The chemical process of uranium separation and purification will itself consume a considerable amount of energy, and in any event the alternative sources will be fully competitive for many years to come.

As coal and oil approach exhaustion, a possible balance may be found by harnessing all the water power of the world and, as far as possible, meeting local power requirements from this source. Where there is a local surplus of water power, the energy could be used to purify fissile material so as to provide power in lands where the alternatives are insufficient. Since fissile material, even in its natural state, is easily transportable, such packaged fuel might solve the power transmission problem.

Water power, tidal and wind power will, no doubt, be developed in many lands to minimize the need for imported fuel, and the world production of industrial alcohol will probably be greatly increased to provide much-needed liquid fuel. Indeed, one of the problems of the Atomic Age will be the provision of small mobile power units, and it may be that part of the answer will be found in the use of nuclear energy via a fuel cell. As far as can be seen at present, nuclear energy will be made available to the people as electricity. Electric rail and road transport will assume a new importance. Britain will become a smokeless zone.

In sum, therefore, man having evolved during a million years has, over the past 250 years, developed a mode of living which is unique in human history. This achievement has thus lasted for only 0·025% of his sojourn here and already it has made great demand on the energy resources of our planet. Unless he is able and willing to match his technology to the unfolding needs of the situation, he has no prescriptive right to a continuance of this latest civilization, and the history of civilizations discloses that discontinuity—decay and rebirth—is the normal method by which successive stages have been reached.

However, to-day man is equipped with a knowledge of natural laws (science) and an ability to harness these laws to his needs (engineering) that were absent in all previous civilizations, and there is good reason for thinking that the present mode of living can be greatly prolonged if he will but use this knowledge and ability aright. The test is one of competence in the political sense to learn how to live together in peace, and in the technical sense to unravel the unsolved relationships in nature and constrain them to serve his ends. The political aspect, though extremely important, and something for which all must share the responsibility, is outside our immediate scope this evening. But the progress of science and technology is the main object for which this Institution exists and, as events are shaping, is one of the two vital matters that will condition the future of the human race.

These twin challenges must be met if civilization is to endure, and members of this Institution are in the front line on one of the battlefields. It is an exciting situation. The need for more and better physicists, chemists and engineers was never more clamant, the results of achievement were never more worth while, and it is extremely important that in our homes and schools the need and the prospect should be fully explained to the rising generation on whose shoulders the responsibility for continuity lies.

We stand before the portals of an epoch. It is the privilege of scientists to unlock one of the doors through which humanity may pass to the enjoyment of a fuller and a freer life for the next 5 000 years.

Let us all see to it that man is worthy of his achievements.

Acknowledgments

This Address contains matter derived from general reading, but the author wishes to acknowledge having read and drawn upon the following recent publications:

Ayres and Scarlott: "Energy Sources—The Wealth of the World" (McGraw-Hill Book Co., 1952).

Palmer Putnam: "Energy in the Future" (Macmillan and Co.).

Report of the U.S. President's Materials Policy Commission.

Abstract No. 1805 R
Jan. 1955

RADIO SECTION: CHAIRMAN'S ADDRESS

By C. W. OATLEY, M.A., M.Sc., Member

"ELECTRICAL CONDUCTION IN NON-METALLIC SOLIDS"

(ABSTRACT *of Address delivered* 13*th October*, 1954.)

In recent years radio engineers have been making increasing use of devices which depend for their action on electrical conduction in non-metallic solids, and the introduction of transistors has made us appreciate the benefits likely to be derived from a study of this subject. Of transistors as such I shall say nothing further here, but I thought it might be useful if I attempted a survey of other devices in which non-metallic conduction plays a part. In most of these devices the crystalline state of the solid is of vital importance, and I shall be speaking chiefly of what takes place inside a crystal which is nearly perfect.

Conditions inside a Non-Metallic Crystal

Let us first consider how electrical conduction might take place in a non-metallic crystal such as sodium chloride for example. In metals there is an abundance of free electrons which are not attached to particular atoms and which can therefore move freely under the influence of an applied electric field, but in sodium chloride each electron is tightly bound to its parent nucleus so that the crystal is normally an insulator. From the results of X-ray analysis we find that the crystal consists of a lattice of positive and negative ions, separated by distances of about $2\cdot 8 \times 10^{-8}$cm. We all know that, in fact, each ion consists of a nucleus surrounded by various groups of electrons, but we do not perhaps always stop to consider the sizes of the components of these ions. Without specifying too exactly what is meant by the radius of a nucleus or of an electron, we may say that the effective radius of each of these particles is of the order of 10^{-13}cm, or about one hundred thousandth part of the distance separating two adjacent ions. Thus the interior of the crystal consists almost entirely of empty space in which the nuclei and the electrons occupy about one million millionth of the total volume. I emphasize this point because I think we sometimes feel instinctively that it must be difficult for an electron to thread its way between the solid particles which make up a crystal. In fact, there is plenty of room.

But if the sizes of the component particles which make up the crystal are insignificant, the electrostatic forces which they exert on each other by virtue of their charges are very large. The electric field which these charges produce will vary enormously from point to point throughout the lattice but is usually much greater than any that we can produce artificially and may easily be as high as $1\cdot 5 \times 10^9$ V/cm.

To sum up, the interior of a crystal of sodium chloride or any other non-metal consists largely of empty space in which very strong electric fields exist. The positively charged nuclei are arranged in a regular lattice, and each is surrounded by groups of electrons in motion, the whole forming an extraordinarily complex dynamical system.

Conditions for Electrical Conductivity to Exist

If an electric current is to flow through a crystal it must be carried by the motion of some of the charged particles about which I have been speaking. Although it is not impossible for complete ions to move through the lattice, I shall not be concerned with conditions under which this happens, so the positions of the positively charged nuclei may be taken to be fixed. The innermost rings of electrons are so tightly bound to their respective nuclei that there is little chance of their breaking away, so they also are unlikely to contribute to electrical conduction. Let us therefore concentrate our attention on the outermost electrons attached to each nucleus, the ones least tightly bound, and consider the conditions under which they exist.

To simplify our problem we consider only motion in one dimension. If we draw any straight-line path through the crystal, the electrical potential will vary periodically along this path because of the periodic structure of the crystal. The forces governing the motion of the outermost electrons will therefore be somewhat analogous to those acting on a number of ball bearings trapped in the hollows of a piece of corrugated iron [Fig. 1(*a*)]. The balls are in motion and therefore have energy, and, in our one-dimensional gravitational analogy, this can be represented by supposing them to oscillate about their lowest positions up to the level of the line AB. Since we are representing conditions in an insulator, the level of AB is below the crests of the corrugated iron so that the balls are tapped in the troughs. If this were not the case the balls would have sufficient energy to surmount the crests; they would be free and we should be representing what happens in a metallic conductor.

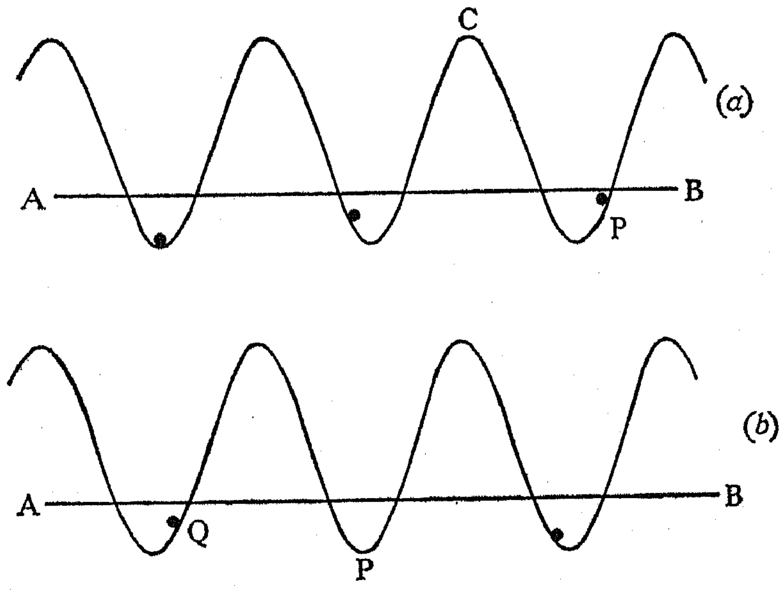


Fig. 1.—Gravitational analogy for the motion of an electron in a non-metallic crystal.

Suppose now that one of the balls, the one at P, say, is given an impulse towards A by some external means, so that it has sufficient energy to surmount the first crest at C. Neglecting friction, it will retain the whole of this energy when it falls into the next trough and will thus be able to climb over the next crest, and so on. In my diagram it looks as though the ball would collide with the one in the next trough, but we must remember that the case of electrons in a crystal is a three-dimensional one

---

Mr. Oatley is in the Department of Engineering, University of Cambridge.

and that the electrons occupy a minute fraction of the total space, so electron-electron collisions would occur very rarely.

If the analogy which I have used gives a true picture of the state of affairs inside an insulating crystal, it is clearly necessary to modify our traditional ideas about the nature of an insulator. We have been accustomed to think of such a substance as one through which an electric current cannot flow or, in terms of electrons, as one through which electrons cannot move freely. We now see that the only hindrance to electronic motion is that the electrons cannot escape from their parent nuclei. If they could be released by some external agency, they could move through an insulating crystal under the action of an applied field as freely as through a metal.

I hope the simple gravitational analogy has been useful in making plausible the idea of electronic motion in an insulator, and I shall use it again to illustrate another aspect of this motion, but I must warn you that it greatly over-simplifies the picture and is faulty in several respects. Electronic motion in crystals can properly be dealt with only by the methods of quantum mechanics, but it turns out that a rigorous discussion of the problem leads to the same conclusion as our simple argument: that an electron once freed can move without difficulty through the crystal. The rigorous discussion also leads to a conclusion which cannot be reached by simple arguments, but which can be illustrated by reference to our simple model. Suppose that, in a particular part of the crystal, one of the electrons has received an impulse and has moved out of the region, leaving a vacancy as at P in Fig. 1(*b*). Then it turns out that an electron from either of the positions adjacent to this vacancy can move into it without difficulty even though its energy was previously insufficient to enable it to surmount the potential hill separating the two troughs. But if an electron from Q, say, moves to P, we may express the fact by saying that the vacancy or hole, as it is termed, has moved from P to Q. This process can go on indefinitely, and so the hole can wander freely through the crystal. Furthermore, since every jump which the hole makes in one direction is in reality caused by an equal jump of an electron in the opposite direction, the net transfer of electricity under the action of an applied field is the same as it would be if the hole had a positive charge equal in magnitude to the negative charge on an electron. Thus it has become common to speak about the motion of positively charged holes, and this motion constitutes a second process by which conduction can take place in a non-metallic crystal. It is by no means obvious, but it turns out that the speed with which a hole moves through the crystal is about the same as that with which an electron would travel through the same crystal under similar conditions of applied field.

It may now be helpful to use another analogy and to liken the conduction of electricity through a non-metallic crystal to that through a gas. Both are normally insulators, but free electrons may be produced in either by ionizing processes which I shall shortly consider and both then conduct as a result of the motion of the free electrons. In the gas the production of free electrons leaves behind positive ions which move under the action of an applied field in a direction opposite to that of the electrons and so constitute a second process of conduction. Similarly, in the crystal, the part of the positive ions is played by the positive holes. However, before pushing this analogy further, it is necessary to say something about irregularities in the crystal.

Irregularities in the Crystal

Hitherto I have spoken about a crystal as though its lattice structure were perfect, but no such crystal exists. In the first place we cannot, in practice, deal with perfectly pure substances, so the lattice will contain foreign atoms with different properties. With existing chemical techniques it is difficult to reduce impurities in most substances below about one part in a million. This means that our crystal is likely to contain about 10^{17} impurity atoms per cubic centimetre. Secondly, even if pure material were available, it would be impossible to grow a perfect crystal. Thermodynamical arguments show that, at room temperature (and still more so at higher temperatures) there will always be a proportion of atoms (or ions) out of place in the lattice. Thirdly, the particles forming the lattice are in continuous irregular oscillation about their mean positions, as a result of their thermal energy. At room temperature the root-mean-square amplitude of these oscillations might be about 10% of the distance between adjacent particles.

For all of these reasons the lattice falls short of perfection, and the periodic curve which I have hitherto used to represent the variation of potential along a straight line in the crystal must be modified accordingly. One result of this can be mentioned at once. Because of the irregularities, a free electron in its passage through the lattice will, from time to time, encounter abnormally high potential hills which it cannot cross and will be deflected by them. In the absence of an external field it will therefore pursue a random zigzag path, much as does a free electron when traversing a gas. If an external field is applied, the motion of the free electron will still be largely random, but there will now be a net drift in the direction in which the field is urging it. This again is analogous to what happens in a gas, and we may borrow the terminology of gaseous conduction and speak of the mobilities of free electrons and holes. Similarly, we may define a recombination coefficient for the union of free electrons and positive holes.

There are so many similarities between conduction in gases and in non-metallic crystals that it may be well to mention some of the differences. In gases the positive ions are always much heavier than the free electrons and move with correspondingly lower velocities. In non-metallic crystals the electrons and positive holes often have nearly the same mobility. Impurities (with which lattice defects must be included) play a much larger part in conduction in crystals than they do in most gaseous discharges. There are several reasons for this. For one thing they are localized in the crystal and cannot be swept away by an electric field if they become charged. Then, again, some impurities in a crystal have their outermost electrons so loosely attached that most of the impurity atoms become ionized as a result of their thermal energy at room temperature, and thus provide a source of free electrons in the crystal. No gas has a low enough ionization potential to act in this way. Similarly, some impurities in a crystal have a sufficiently great affinity for electrons to be able to capture them from the lattic at room temperature and thus provide a source of free positive holes. It follows that, when the carriers in a crystal are provided by impurities they may be preponderantly of either sign, depending on the nature of the impurities. In a gas the positive ions and electrons are generally present in comparable numbers. These are very real differences, and many of the effects which can be produced in crystals have no counterpart in gases.

Production of Free Carriers by Thermal Means

So far, the presence of free electrons and holes in a crystal has been assumed, but little has been said about the ways in which they may be produced. The basic necessity is for excess energy to be communicated to the most loosely bound electrons so that they may be detached from their parent nuclei, and there are three general ways in which this may be done. The temperature of the crystal may be raised to the point where thermal energy of the lattice is sufficient to release some of the electrons; swiftly-moving particles such as electrons or α-rays may be allowed to fall on the crystal; or the additional energy may be provided in the form of photons or quanta of radiant energy. It will be convenient to examine these methods one by one and, at the same time, to consider what practical devices make use of them.

At room temperature the electrons in a crystal will be in equilibrium with the lattice, and calculation shows that a certain fraction of them will be free. However, unless the energy required to detach an electron from its nucleus is considerably less than one electron-volt, the number of free electrons will be too small to produce appreciable conductivity. There are very few pure substances for which the ionization energy is as small as this—germanium is one of them—but in many impure substances the ionization energy of the impurities is quite low. Such substances are semi-conductors. When they are heated the number of free carriers present in them increases because greater thermal energy is available in the lattice and so the conductivity of the material increases. Practical devices making use of this effect are known as thermistors and find application in many radio circuits.

A much more important class of device making use of the properties of semi-conductors is to be found in the wide range of non-linear conductors which are now in production. These include selenium and copper-oxide rectifiers, point-contact silicon and germanium diodes, and the silicon-carbide resistors which are used in protective circuits.

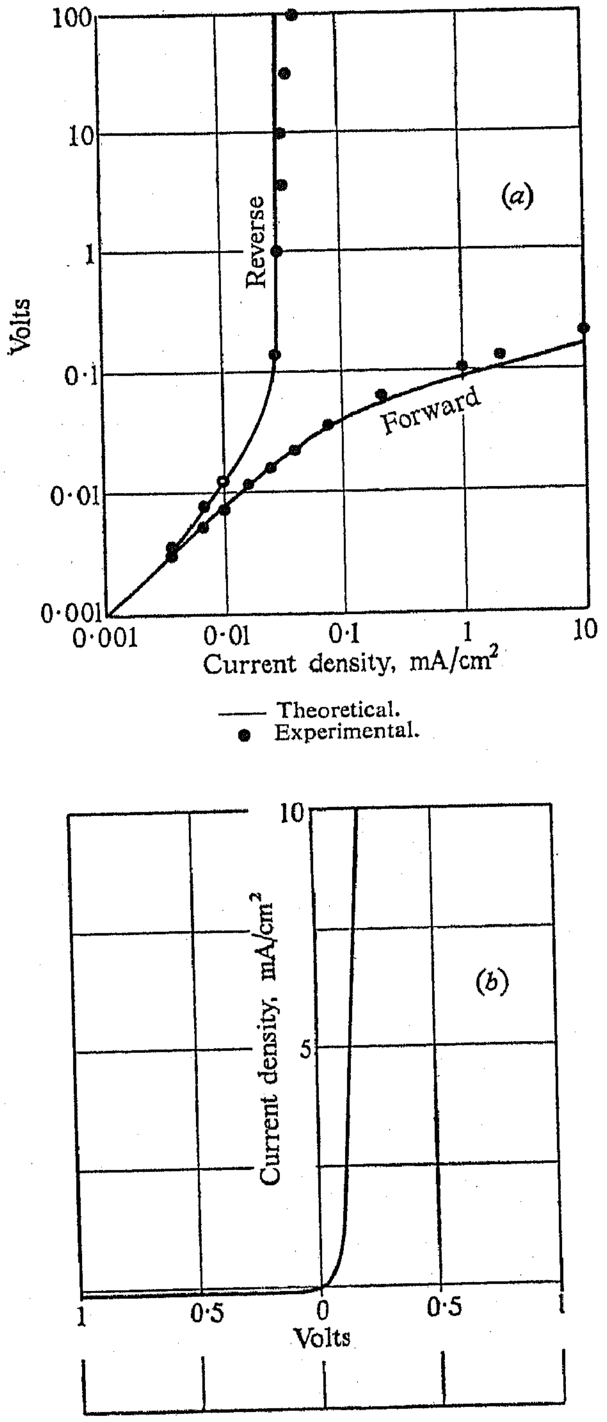


Fig. 2.—Theoretical and experimental curves for a germanium junction diode.

Most of these non-linear conductors were developed before the theory of semi-conduction was properly understood, and in many of them the conditions are so complex that even now it is difficult to apply the theory to them. In most of them we are dealing not with a single crystal but with a micro-crystalline mass of variable and uncertain composition. Nevertheless, many of the properties of these rectifiers can now be satisfactorily explained, and there is little doubt that our knowledge of the theory will lead to improved performance in future. In contrast, we have recently witnessed the introduction of silicon and germanium junction diodes. In these the composition of the material is most carefully controlled and the theory is understood in considerable detail. In Fig. 2(*a*) is shown a comparison between theory and experiment for the current/voltage characteristics of a germanium junction diode. These curves are based on data obtained in the Bell Telephone Laboratories[1] and furnish a striking illustration of the extent to which the theoretical behaviour of these devices is understood.

In Fig. 2(*b*), the experimental curve is plotted in a more conventional manner to emphasize the remarkable properties of these interesting rectifiers. More recently, figures have been published for silicon junction diodes at room temperature, which indicate that reverse currents as low as 10^{-10}amp and rectification ratios as high as 10^8 at 1 volt can be attained.

Production of Free Carriers by the Absorption of Radiation or by the Impact of Swiftly Moving Particles

A non-metallic insulating crystal may become conducting if, as a result of the absorption of radiation (ultra-violet, visible or infra-red), free electrons or positive holes are produced in it. The selenium cell, which has been known for many years, provides an example of this mechanism. In recent times it has been supplanted as a means of detecting and measuring light by other devices which are more stable and more sensitive. However, since the war, cells similar in principle to the early selenium cell have been developed in which the active material is a layer of suitably treated lead sulphide, lead selenide or lead telluride. These cells make excellent detectors of radiation over a useful part of the infra-red spectrum and are about one hundred times as sensitive as any other known detectors. The active layers in these cells are polycrystalline, and it seems certain that their action depends on effects taking place at the grain boundaries.[2]

A more recent application of the photoconductive effect is its use in a new type of television camera tube, one form of which is termed the Videcon.[3] For some purposes, this tube has advantages over the more commonly used types.

I turn next to the phosphors used in the screens of cathode-ray tubes. Usually a phosphor consists of a crystalline powder (e g. zinc sulphide) to which has been added a small percentage of activator (e.g. copper) in such a way that the activator atoms become incorporated in the lattice of the parent crystals. From measurements of the efficiency of such screens and from the colour of the light emitted (which is characteristic of the activator), it is clear that most of the energy of the incident electrons is communicated to the activator centres; yet few of the electrons collide directly with these centres.

To explain this energy exchange we assume that the initial process involves the creation of free holes and electrons, and that the activator centres become ionized by combination with the holes thus formed. Finally, the ionized activator centres unite with the free electrons and the excess energy thus liberated is radiated as light.

Although there is much evidence to show that this mechanism operates in many phosphors, it is inadequate to explain the behaviour of those which exhibit pronounced afterglow. In this case it is necessary to postulate the existence of irregularities in the crystal which attract electrons sufficiently strongly to immobilize them. Such irregularities are termed traps. The mechanism described above will now be complicated by the trapping of some of the free electrons, which cannot then combine with the ionized activator centres until they are released from the traps. This release is usually brought about by the thermal energy of the lattice, but may take appreciable time, so the emission of light decays gradually.

The existence of electron traps provides a simple explanation of an experimental observation which would otherwise be puzzling. In the whole of what I have been saying on this subject it has been assumed that a non-metallic crystal will become conducting if sufficient energy is communicated to the valence electrons by incident radiation or swift particles. However, if we evaporate metallic electrodes on to two opposite sides of such a crystal and connect them in a circuit with a battery and galvanometer, it is often found that no appreciable current flows when the crystal is irradiated. In explaining this, the first point to be made is that the incident radiation or particles will usually penetrate the crystal to only a very small depth. By hypothesis they are giving up energy to the crystal, so they will be strongly absorbed. In general, the electrons and holes formed in the region which is penetrated will have to travel across the rest of the crystal to the metal electrodes, and if the electrons become trapped in appreciable numbers, they will set up a space charge which completely neutralizes the applied field and causes the flow of current to cease.

That this actually happens is shown by some experiments recently carried out by McKay,[4] in which a diamond was bombarded with a pulse of high-energy electrons lasting only 5 microsec, and the current flowing across the diamond round a separate circuit was measured during this interval. It was found that the current decayed rapidly from its initial value, the rate of decay being smaller the greater the field applied to the crystal. This is what we should expect, since, with larger values of the applied field, it will take longer for the necessary space charge to build up. However, neutralization of the applied field was virtually complete after a few microseconds, even when the voltage applied across the crystal was as high as 500 volts.

The conditions are quite different if the crystal is in the form of a very thin slice so that the incident radiation or particle can pass right through it without too much loss of energy. Under these circumstances, free electrons and holes are continually being produced throughout the crystal and any electrons which are trapped will rapidly be neutralized by recombination with holes. There will then be no appreciable space-charge and the conductivity of the crystal can readily be observed. The conductivity produced in this way by a beam of electrons has been made the basis of operation of a picture storage tube known as the Graphechon.[5]

The Present Position with regard to Non-Metallic Conduction

Summing up the present position with regard to non-metallic conduction, we may say that the basic phenomena are beginning to be well understood but that in nearly every case of practical interest there are complicating effects which have not yet been fully elucidated. For example, our knowledge of what takes place at the junction of a metal and a semi-conductor is still very far from complete.

On the experimental side I have attempted to show the immense effect that impurities can have even if they are present to the extent of less than one part in a million, and by these standards very few pure chemical compounds have ever been prepared. When the problems of purification have been overcome, the task of growing single crystals of the required materials and of introducing controlled amounts of known impurities into the lattice will remain. Finally we shall have to learn how to use the new materials in new practical devices.

It is apparent, therefore, that we are only at the beginning of a very long story. The position is perhaps comparable with that of the thermionic valve thirty years ago. In 1924 the laws of thermionic emission were understood and we knew how to make thoriated-tungsten and oxide-coated cathodes. Vacuum technique had made great progress and the laws of electronic motion and of space charge were known. But only diodes and triodes had emerged as practical devices.

References

(1) Goucher, F. S., Pearson, G. L., Sparks, M., Teal, G. K., and Shockley, W.: "Theory and Experiment for a Germanium *p-n* Junction," *Physical Review*, 1951, **81,** p. 637.

(2) Smith, R. A.: "Infra-Red Photo-Conductors," *Advances in Physics*, 1953, **2,** p. 321.

(3) Weimer, P. K., Forgue, S. V., and Goodrich, R. R.: "The Videcon—Photoconductive Camera Tube," *Electronics*, 1950, **23,** p. 70.

(4) McKay, K. G.: "Electron Bombardment Conductivity in Diamond," *Physical Review*, 1948, **74,** p. 1606.

(5) Pensak, L.: "The Graphechon—A Picture Storage Tube," *R.C.A. Review*, 1949, **10,** p. 59.

621.372.2: 621.315.61

Paper No. 1699 R
Jan. 1955

THE PROPERTIES OF ARTIFICIAL DIELECTRICS AT CENTIMETRE WAVELENGTHS

By JOHN BROWN, M.A., Associate Member, and WILLIS JACKSON, D.Sc., D.Phil., F.R.S., Member.

(*The paper was first received* 28*th January, and in revised form* 11*th May,* 1954.)

SUMMARY

The properties of an artificial dielectric at centimetre wavelengths are compared with those of a solid material, and the possible definitions of the refractive index and the wave impedance are examined. It is concluded that the most appropriate definition of the latter is in terms of the reflection coefficient at an interface between the artificial dielectric and free space. The reactive fields which are set up near such an interface are discussed qualitatively, and a number of ways in which they may be represented are given. The paper concludes with suggestions for further applications of artificial dielectrics.

(1) INTRODUCTION

The increasing use of optical techniques at microwave frequencies has led to a search for media which refract electromagnetic waves in the same way as glass refracts light waves. An obvious example is a solid dielectric such as polythene or polystyrene, but the cost and weight of these materials make them unsuitable if large volumes are required, as, for example, in the construction of lens aerials. A convenient substitute was described by Kock[1] in 1944 and consists of what is, in effect, a large-scale model of a solid dielectric, the molecules of which are replaced by highly conducting elements. Such media can be made to have a very low density and they reproduce the essential properties of solid dielectrics. They are frequently referred to as "artificial dielectrics," a term which may conveniently be generalized to include any regular array of conducting elements which refracts electromagnetic waves.

At low frequencies, artificial dielectrics show an effective permittivity which can be calculated from the change in capacitance caused by inserting a slab of the medium between the plates of a capacitor. Results for a selection of the more useful media have already been reported.[2] Theoretical estimates of the relative permittivity may be obtained by using the Lorentz theory for non-polar dielectrics. This theory has a limited range of application, but reasonable agreement with experimental results is obtained if allowance is made for the short-range interaction forces between nearby conductors.

Certain artificial dielectrics have a low-frequency permeability which differs from the free-space value. No measurements of these values have been reported, but theoretical values may be obtained by a method similar to that used for the permittivity. In any case the most useful artificial dielectrics do have the same low-frequency permeability as free space.

When artificial dielectrics are used at high frequencies there is no guarantee that they will behave in the same way as would a continuous medium with the same low-frequency values of permittivity and permeability. An obvious possibility is that the properties of the medium may be frequency-dependent, i.e. that it is dispersive. A more fundamental consideration is what meaning can be attached to permittivity and permeability at very high frequencies. Since electric and magnetic fields are then inextricably mixed, it becomes impossible to make independent measurements of the permittivity and the permeability. It is, therefore, desirable to make a thorough investigation of the behaviour of artificial dielectrics at high frequencies, and this is the subject of a research programme at present being carried out in the Department of Electrical Engineering at Imperial College. The first results of this are presented in the following papers by M. M. Z. El-Kharadly[3] and R. I. Primich,[4] to which the present account of the general properties of artificial dielectrics at high frequencies is regarded as a desirable introduction. A clear picture of the corresponding properties of solid materials is necessary for this purpose and these are summarized in the next Section.

(2) WAVE PROPAGATION IN HOMOGENEOUS SOLID MATERIALS AT HIGH FREQUENCIES

The electromagnetic properties of a homogeneous solid material are described by three quantities: namely, the permittivity, $\epsilon = \epsilon_0\epsilon_r$; the permeability, $\mu = \mu_0\mu_r$; and the conductivity, σ; where ϵ_r and μ_r denote the relative permittivity and permeability respectively. It will suffice to restrict the discussion to perfect insulators, for which σ is zero, and to assume that ϵ and μ are independent of frequency. The simplest and most common type of propagation in a solid material is a plane wave, for which the field distribution is well known. If the directions of the three mutually perpendicular vectors, the electric field, the magnetic field and the direction of propagation are taken as the x-, y- and z-axes respectively, then

$$E_x = E_0 \exp j(\omega t - \beta z) : H_y = (E_0/Z) \exp j(\omega t - \beta z) \quad . \quad (1)$$

all the other field components being zero. The various parameters are E_0, the arbitrary amplitude constant; ω, angular frequency; β, phase-change coefficient; and Z, wave impedance.

The last two are related to ϵ and μ by

$$\beta = \omega(\epsilon\mu)^{\frac{1}{2}} \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

$$Z = (\mu/\epsilon)^{\frac{1}{2}} \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

The primary consideration is what can be deduced about the properties of the medium from measurements on the plane wave. The only quantities which can be directly measured are the amplitudes and phases of the electric and magnetic fields. From the rate of change of phase of either field the phase-change coefficient can be immediately obtained, while the wave impedance can be calculated as the ratio of the magnitudes of the fields at any point. The essential feature is that β and Z can be measured independently. There is, however, no way by which ϵ and μ may be measured independently, and indeed the only way to obtain ϵ and μ is to calculate them from eqns. (2) and (3) using measured values of β and Z.

Written contributions on papers published without being read at meetings are invited for consideration with a view to publication.

Mr. Brown is in the Department of Electrical Engineering at Imperial College, University of London.

Dr. Willis Jackson, who was formerly Professor of Electrical Engineering at Imperial College, University of London, is now with the Metropolitan-Vickers Electrical Co., Ltd.

A similar conclusion can be reached for any type of propagation in the solid material—when inserted, for example, in a waveguide—and it is worth recalling that the well-established methods for measuring ϵ by the use of cavity resonators rest on the assumption that μ for the material being investigated has the free-space value, μ_0.

The above conclusions require qualification in one respect. In a cavity resonator there are regions, of dimensions small compared with the wavelength, in which the value of the magnetic field is low, and a measurement of the permittivity of a solid material may be made by placing a very small sample in such a region. There are other regions in which the electric field is low and the permeability can be obtained in a similar manner. This technique cannot be applied to artificial dielectrics, however, since the sample must include several conducting elements before it displays the same properties as a large sample. The use of samples whose dimensions are small compared with the wavelength is therefore excluded.

It appears, then, that the basic parameters of a material at high frequencies are not ϵ and μ but β and Z. This has long been realized by workers in optics, who use the refractive index of a material as its fundamental constant. Any of the definitions of refractive index lead to the equation

$$\beta = n\beta_0 \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

where β_0 is the phase-change coefficient for a plane wave in free space. The refractive index, n, is more convenient to use than β, to which it is directly related by the above equation. The second parameter, Z, is invariably equal to Z_0/n for the materials used in optics, Z_0 being the wave impedance of a plane wave in free space. The wave impedance is therefore not usually considered in optical problems, but it must be retained in the radio case.

Although the method suggested for the measurement of Z is a feasible one in principle, it presents practical difficulties, and a more convenient method is to measure the amplitude reflection-coefficient, ρ, when a plane wave is incident from free space normally onto a semi-infinite block of the material and then to calculate Z from the equation

$$Z = Z_0\frac{1 - |\rho|}{1 + |\rho|} \quad . \quad . \quad . \quad . \quad . \quad . \quad (5)$$

(3) PROPAGATION IN ARTIFICIAL DIELECTRICS

(3.1) Definition of Refractive Index

The precise meaning of the quantities n and Z for an artificial dielectric can only be determined by examining in detail the properties of a wave being propagated through the medium. For simplicity, the strip type of metallic-delay dielectric, Fig. 1, is considered and the strips are assumed to be placed in free space. The directions of the axes of the strips and of propagation are taken as the y- and z-axes respectively, and the only field components which are excited are E_x, E_z and H_y. These components do not vary in the y-direction and so are functions of x and z only. Any component, E_x, for example, may be written in the form

$$E_x = \phi(x,z)\exp(-j\beta z) \quad . \quad . \quad . \quad . \quad . \quad (6)$$

where $\phi(x,z)$ is a doubly periodic function with the property

$$\phi(x + pb, z + qa) = \phi(x,z) \quad . \quad . \quad . \quad . \quad (7)$$

a, b being the lattice spacings shown in Fig. 1, and p, q being integers.

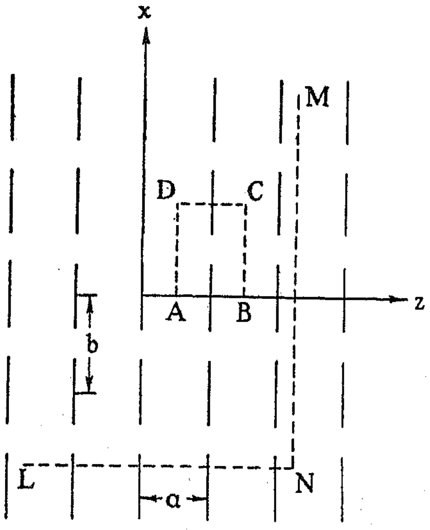


Fig. 1.—Cross-section of strip-type delay dielectric.

Expressions similar to that in eqn. (6) occur in the band theory of solids[5] and the methods developed for this subject may be used to analyse the performance of artificial dielectrics.[6]

Eqns. (6) and (7) state mathematically the condition that the fields in any cell such as ABCD in Fig. 1 differ from one another only by the phase factor, exp $(-j\beta z)$. The phase difference between any two corresponding points L and M is βqa, where qa is the distance LN. If the artificial dielectric is replaced by a solid material with the same phase-change coefficient the phase difference between L and M is again βqa. The refractive index of the solid is β/β_0 and this suggests that the same value be taken for the artificial dielectric. This is obviously permissible if lengths consisting of exact multiples of a only are considered.

The phase varies between the points L and N as shown in Fig. 2. Instead of the linear change, shown by the broken line, which applies for a solid material, the variation is oscillatory, the period of the oscillation being a. The amplitude of the oscillation depends on the function $\phi(x,z)$ and indeed the difference between the curve and the straight line is just the phase of $\phi(x,z)$. The artificial dielectric can be said to have a definite value of refractive index only if the departures of the curve from the straight line can be ignored. The value taken for the refractive index is obtained by averaging the rate of change of phase over a distance equal to the spacing a.

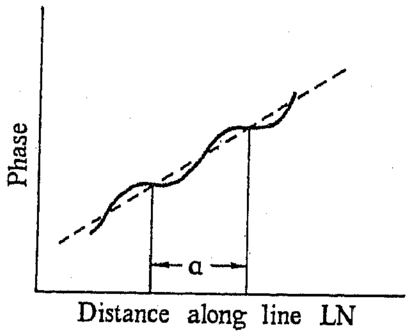


Fig. 2.—Variation of phase in artificial and solid dielectrics.

——— Artificial.
— — — Solid.

(3.2) Definition of Wave Impedance

The problem of defining a suitable value for the wave impedance is not so simple. If the original definition in terms of the amplitudes of the electric and magnetic field strengths is retained, the wave impedance is a function of position. A definite value could be obtained by averaging either over the volume of the cell, or over a surface perpendicular to the direction of propagation. However, this merely transfers the problem elsewhere, for such an impedance could be measured only by plotting the complete field distribution throughout the cell and then averaging in the appropriate way. Further, it is far from clear what significance this quantity would have: for example eqn. (5) is most unlikely to be valid except in very simple cases. An averaging process is not therefore to be recommended since there is no simple method of measuring the quantity so defined, nor does it have any property of particular significance.

An alternative procedure is to take eqn. (5) as the definition of the wave impedance. This has much to recommend it since the wave impedance appears in calculations only when junctions between different media are involved. Further, the wave impedance, defined in this way, may be measured by the method suggested for solid materials. The values given by El-Kharadly[3] are for the wave impedance in this sense.

(3.3) The Reactive Fields near an Interface

There is another important consideration which must not be overlooked. The fields, set up when a plane wave is incident upon a semi-infinite block of an artificial dielectric, include in general a reactive contribution limited to the immediate vicinity of the interface. In the most widely studied example of this, the artificial dielectric consists of a set of parallel conducting plates and the reactive field causes changes in the phases of the reflected plane wave and of the wave transmitted into the dielectric.[7] These phase-shifts must be specified before the properties of the artificial dielectric are fully known.

The strip dielectric, Fig. 1, may be used to illustrate the way in which this interface effect arises. A previous paper[8] has given a simple equivalent circuit, consisting of a transmission line loaded by shunt reactances at equal intervals, a, as shown in Fig. 3(b). This circuit is valid if the spacing a exceeds $0 \cdot 75b$. The wave impedance, as defined by eqn. (5), is in this case equal to the iterative impedance of a filter made up from a succession of sections such as LM in Fig. 3(b). Further, there are no phase-shifts if the effective interface is taken as a plane distant $a/2$ in front of the first plane of strips, i.e. QR in Fig. 3(a).

When the spacing a becomes less than $0 \cdot 75b$ the reactive fields, which are associated with each plane of strips and represented in Fig. 3(b) by the reactances, X, are no longer completely isolated. A reactive coupling now exists between successive planes of strips, and the equivalent circuit takes the form of Fig. 3(c), the series reactances, X', arising from this reactive coupling.[9] The reactive fields at the interface come from the element NP, one end of which has been left free. There is no obvious method of terminating this element to give a correct representation of the interface effect. A preliminary analysis of the behaviour at the interface, which is still in progress, shows that additional modes of propagation must occur within the artificial dielectric. These are evanescent in the cases of immediate interest. So far no theoretical values for the phase-shifts at the interface have been obtained.

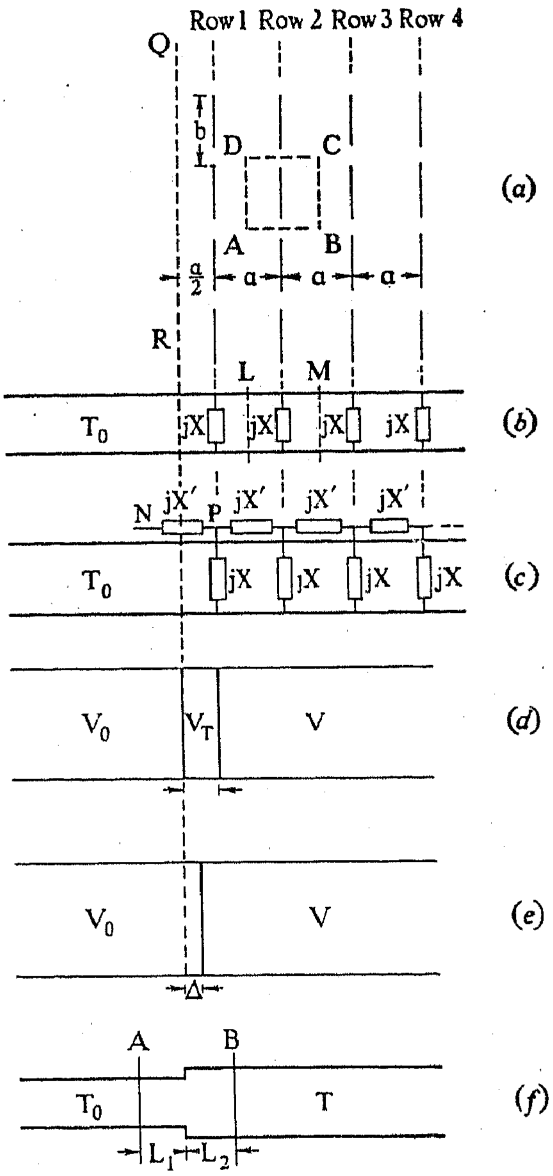


Fig. 3.—Representations of the artificial dielectrics.

(a) Free-space interface.
(b) Simple equivalent circuit valid if $a > 0 \cdot 75b$.
(c) Equivalent circuit if $a < 0 \cdot 75b$.
(d) and (e) Alternative representations in terms of solid materials.
(f) General equivalent circuit.
T_0 is a transmission line with phase-change coefficient β_0 and characteristic impedance Z_0.
T is a transmission line with phase-change coefficient β and a characteristic impedance equal to the wave impedance of the artificial dielectric as defined by eqn. (5).
V_0 represents free space.
V_t represents a transition layer of unknown parameters.
V represents a solid material with the same phase-change coefficient and wave impedance [in the sense of eqn. (5)] as the artificial dielectric.

An alternative explanation of the interface effect can be obtained from the Lorentz theory for non-polar dielectrics.[2] The dipole moment induced in each conductor depends on the effective field, i.e. the sum of the incident field, E_0, and an interaction field, E_1, caused by the dipoles induced in the other conductors. In an artificial dielectric of infinite extent, E_1 has the same magnitude at each conductor, but this cannot be true if an interface exists as in Fig. 3(a). The general behaviour near the interface can be deduced by making the simplifying assumption that E_1 arises primarily from the elements immediately adjacent to that being considered. The reaction field may then be assumed of constant magnitude for any element in rows 2–3 . . . in Fig. 3(a), but it must be different for the elements in row 1 since there is no row to the left of this. The dipole elements induced in the elements of the first row differ from those in the other rows and there must, therefore, be a transition region between free space and the artificial dielectric. This region may be represented in many ways. A simple one is to represent a region such as ABCD of Fig. 3(a), which contains a line dipole of moment m per unit length in the y-direction, by a continuous material with polarization mA where A is the area of the region ABCD. The structure of Fig. 3(a) is then equivalent to that in Fig. 3(d) in which a transition layer is interposed between free space and the continuous medium corresponding to the main bulk of the artificial dielectric. An alternative is to make the polarization in the transition layer have the same value as in the remainder of the dielectric and to change its volume so that the total dipole moment in the transition layer is the same as for the first row of strips. This gives Fig. 3(e) which shows a shift Δ in the position of the interface.

The argument of the previous paragraph is far from rigorous since only changes in the electric polarization have been considered. Further, the interaction field acting on an element may arise from elements other than those immediately adjacent to it. The technique may be extended to allow for these other considerations by replacing the single transition layer of Fig. 3(*d*) by a number of layers with varying values of refractive index and wave impedance. In principle, the representation of the artificial dielectric may be made as accurate as may be desired but at the expense of introducing additional parameters. Since there is as yet no theory from which these may be calculated, a technique has to be developed by which they may be obtained experimentally. El-Kharadly[3] has done this for the representation given by Fig. 3(*e*).

(3.4) Other Representations of Artificial Dielectrics

The general representation outlined at the end of the previous Section is more detailed than is required for most practical problems. It provides a method by which the average field over any cell may be deduced from an equivalent assembly of continuous media. Similar suggestions have been previously put forward by Corkum[10] and Susskind[11] but neither author has given a detailed account of how the parameters in their representation could be calculated or measured.

The most important practical problem associated with artificial dielectrics is the determination of the phase-shift caused by a slab of the medium and of the amount of power reflected from it. This does not require a detailed knowledge of the fields near an interface but only a measure of the phase changes in the principal transmitted and reflected waves. It is sufficient for this to specify the transmission and reflection coefficients with respect to a chosen interface such as QR in Fig. 3(*a*): complex values of these coefficients indicate the presence of phase changes. When two or more interfaces are involved, the overall reflection and transmission properties can be calculated with the help of general formulae. This approach has been used by Primich[4] in a study of the interface effects for a medium formed from parallel, perfectly conducting, plane surfaces.

While the specification of reflection and transmission coefficients gives all the necessary information, it lacks the tangibility of an equivalent circuit. It is easy to convert the one into the other, however, and the possible equivalent circuits are identical in form to those which have been studied in connection with waveguide junctions.[12] The most appropriate is shown in Fig. 3(*f*): transmission lines T_0 and T represent free space and the artificial dielectric, respectively. The impedance of T is equal to the wave impedance as defined by eqn. (5). Distances on the free-space side of the interface QR, Fig. 3(*a*), correspond to distances measured from A in Fig. 3(*f*) and distances in the medium from QR correspond to distances from B. The additional lengths, L_1 and L_2, of the transmission lines inserted between A and B are determined by the phases of the reflection and transmission coefficients. The circuit has the advantage that structures involving more than one interface can be analysed using only transmission-line formulae. There is no fundamental difference between the use of reflection and transmission coefficients and the equivalent circuit, and a choice between the two rests on personal preference.

(4) DEPENDENCE OF THE REFRACTIVE INDEX AND THE WAVE IMPEDANCE ON FREQUENCY

Artificial dielectrics generally commence to show dispersion when the wavelength of operation is less than ten times the largest cell dimension. A number of simple formulae, which predict the variation of the refractive index and the wave impedance with frequency, are available for certain of the more common artificial dielectrics. A discussion of these, together with experimental data, is given by El-Kharadly.[3]

One point of considerable interest is whether eqns. (2) and (3) apply to artificial dielectrics in the low-frequency range for which ϵ and μ can be measured independently. Physically this seems to be very probable since a "solid" material is no more solid than an artificial one if sufficiently small volumes are examined. It is, therefore, satisfying to find that a formal justification of the validity of eqn. (2) can be deduced directly from Maxwell's equations, as shown in Section 7.

The question of whether eqn. (3) is valid for artificial dielectrics is more difficult to answer. No general analysis of this has yet been undertaken, but in a number of particular examples the wave impedance as defined by eqn. (5) does satisfy eqn. (3) in the frequency range for which there is no dispersion. Further, no exception to this result has yet been found.

If the analysis given in Section 7 is carried to the next stage of approximation it is found that

$$n = n_0 + \text{a constant times } \beta^2 \quad . \quad . \quad . \quad . \quad (8)$$

n_0 being the value of the refractive index for very low frequencies. The rate of change of n with frequency is therefore zero at zero frequency.

(5) POSSIBLE DEVELOPMENTS

The only practical use to which artificial dielectrics have so far been put is in the construction of microwave lenses. A considerable flexibility in the properties of artificial dielectrics arises from the wide choice of conductor sizes and spacings which is available. A fuller knowledge of the microwave properties may therefore lead to further applications, one such being a method of rapidly scanning a radiated beam by applying frequency modulation to the transmitter and then passing the beam through a highly dispersive prism. Artificial dielectrics are well suited for the prism material and an investigation of this application is now in progress.

A general aspect of artificial dielectrics which has not yet been considered concerns their behaviour when the wavelength is less than the lattice spacings. There is then a close connection with the subject of X-ray crystallography andit may be possible to obtain useful information by making artificial dielectrics as models of particular crystals on which to study the effects of lattice imperfections.

(6) REFERENCES

(1) Kock, W. E.: "Metallic Delay Lenses," *Bell System Technical Journal*, 1948, **27**, p. 58.

(2) El-Kharadly, M. M. Z., and Jackson, W.: "The Properties of Artificial Dielectrics comprising Arrays of Conducting Elements," *Proceedings I.E.E.* (Paper No. 1472 R, July 1953), **100**, Part III, p. 199.

(3) El-Kharadly, M. M. Z.: "Some Experiments on Artificial Dielectrics at Centimetre Wavelengths" (see page 17).

(4) Primich, R. I.: "A General Experimental Method to Determine the Properties of Artificial Media at Centimetre Wavelengths, applied to an Array of Parallel Metallic Plates" (see page 26).

(5) Seitz, F.: "Modern Theory of Solids (McGraw-Hill, 1940).

(6) Bennett, H. S.: "Electromagnetic Transmission Characteristics of the Two-dimensional Lattice Medium," *Journal of Applied Physics*, 1953, **24**, p. 785.

(7) Carlson, J. F., and Heins, A. E.: "Reflection of an Electromagnetic Wave by an Infinite Set of Plates," *Quarterly of Applied Mathematics*, 1947, **4**, p. 313.

(8) Brown, J.: "The Design of Metallic Delay Dielectrics," *Proceedings I.E.E.* (Paper No. 915 R, January 1950), **97**, Part III, p. 45.

(9) COHN, S. B.: "Analysis of the Metal Strip Delay Structure," *Journal of Applied Physics*, 1949, **20**, p. 257.

(10) CORKUM, R. W.: "Isotropic Artificial Dielectrics," *Proceedings of the Institute of Radio Engineers*, 1952, **40**, p. 574.

(11) SUSSKIND, C.: "Obstacle Type Artificial Dielectrics for Microwaves," *Journal of the British Institution of Radio Engineers*, 1952, **12**, p. 49.

(12) MARCUVITZ, N.: "Waveguide Handbook" (McGraw-Hill, 1951).

(13) STEVENSON, A. F.: "Solution of Electromagnetic Scattering Problems as Power Series in the Ratio (Dimension of Scatterer)/Wavelength," *Journal of Applied Physics*, 1953, **24**, p. 1134.

(7) APPENDIX

(7.1) The Refractive Index of an Artificial Dielectric at Low Frequencies

Consider an artificial dielectric formed by a cuboidal array of isolated conductors with lattice spacings a, b and c in the directions of the x-, y- and z-axes respectively. Only propagation in the direction of the z-axis is considered and the field distribution is assumed to be such that conductors may be inserted in the planes $x = \pm a/2$, and magnetic walls in the planes $y = \pm b/2$, without disturbing the fields. This assumption is permissible for delay-type dielectrics. A cross-section of a basic cell is shown in Fig. 4.

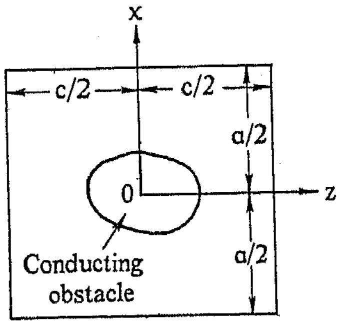


Fig. 4.—Cross-section in the plane $y = 0$ of the basic cell of a delay-type artificial dielectric.

Let $\boldsymbol{E}$ and $\boldsymbol{H}$ be the electric and magnetic fields respectively. The time-dependent factor is taken as $\exp(j\omega t)$ and will be suppressed throughout. The fields must satisfy Maxwell's equations:

$$\text{curl}\,\boldsymbol{E} = -j\omega\mu_0\boldsymbol{H} \quad . \quad . \quad . \quad . \quad . \quad (9)$$

$$\text{curl}\,\boldsymbol{H} = j\omega\epsilon_0\boldsymbol{E} \quad . \quad . \quad . \quad . \quad . \quad . \quad (10)$$

and also the boundary conditions

$$\boldsymbol{i}_\wedge\boldsymbol{E} = \boldsymbol{i}.\boldsymbol{H} = 0 \text{ when } x = \pm a/2 \quad . \quad . \quad . \quad (11)$$

$$\boldsymbol{j}.\boldsymbol{E} = \boldsymbol{j}_\wedge\boldsymbol{H} = 0 \text{ when } y = \pm b/2 \quad . \quad . \quad . \quad (12)$$

$$\boldsymbol{E}(x, y, c/2) = \exp(-j\beta c)\boldsymbol{E}(x, y, -c/2) \quad . \quad . \quad (13)$$

$$\boldsymbol{H}(x, y, c/2) = \exp(-j\beta c)\boldsymbol{H}(x, y, -c/2) \quad . \quad . \quad (14)$$

where $\boldsymbol{i}$ and $\boldsymbol{j}$ are unit vectors in the directions of the x- and y-axes respectively.

Eqns. (11) and (12) arise from the assumption that conductors and magnetic walls can be inserted as described above. Eqns. (13) and (14) express the condition that a wave is being propagated in the direction of the z-axis with a phase-change coefficient β. A method for calculating β when the spacings a, b and c are small compared with the free-space wavelength will now be described.

The fields $\boldsymbol{E}$ and $\boldsymbol{H}$ are expanded in the series

$$\boldsymbol{E} = \boldsymbol{E}_0 + (j\beta_0 c)\boldsymbol{E}_1 + (j\beta_0 c)^2\boldsymbol{E}_2 \quad . \quad . \quad . \quad (15)$$

$$\boldsymbol{H} = \boldsymbol{H}_0 + (j\beta_0 c)\boldsymbol{H}_1 + (j\beta_0 c)^2\boldsymbol{H}_2 \quad . \quad . \quad . \quad (16)$$

in which the terms $\boldsymbol{E}_1$ and $\boldsymbol{H}_1$ are independent of the free-space phase-change coefficient, β_0. Expansions of this type were suggested by Stevenson in a study of diffraction problems.[13] At a sufficiently low frequency $\beta_0 c$ becomes very much less than unity and the expansions reduce to their first terms. Separate equations for the various terms can be obtained by substituting the series in each of eqns. (9) to (14) and then equating like powers of $j\beta_0 c$ on either side. For the present purpose it is sufficient to obtain the equations for $\boldsymbol{E}_0$ and $\boldsymbol{H}_0$: the above procedure gives from eqns. (9) and (10)

$$\text{curl}\,\boldsymbol{E}_0 = \text{curl}\,\boldsymbol{H}_0 = 0 \quad . \quad . \quad . \quad . \quad (17)$$

which shows that $\boldsymbol{E}_0$ and $\boldsymbol{H}_0$ are both static fields. They may therefore be written

$$\boldsymbol{E}_0 = -\,\text{grad}\,\phi \quad . \quad . \quad . \quad . \quad . \quad (18)$$

$$\boldsymbol{H}_0 = -\,\text{grad}\,\psi \quad . \quad . \quad . \quad . \quad . \quad (19)$$

where ϕ is an electrostatic potential and ψ is a magnetostatic potential.

Eqns. (11) and (12) are also valid if $\boldsymbol{E}$ and $\boldsymbol{H}$ are replaced by $\boldsymbol{E}_0$ and $\boldsymbol{H}_0$. Eqns. (13) and (14) are a little more complicated because of the factor $\exp(-j\beta c)$. Since β equals $n\beta_0$ it is of the same order as β_0 and so $\exp(-j\beta c)$ must be expanded in powers of $j\beta c$ before corresponding terms on the two sides of the equation are made equal. It is easily shown that

$$\boldsymbol{E}_0(x, y, c/2) = \boldsymbol{E}_0(x, y, -c/2) \quad . \quad . \quad . \quad (20)$$

$$\boldsymbol{H}_0(x, y, c/2) = \boldsymbol{H}_0(x, y, -c/2) \quad . \quad . \quad . \quad (21)$$

An examination of eqns. (11), (12), (18) and (20) shows that $\boldsymbol{E}_0$ is the electrostatic field established when a constant potential difference is applied between the conductors in the planes $x = \pm a/2$. Similarly, $\boldsymbol{H}_0$ is found to be the magnetic field which is set up by a current flowing along these conductors in the z-direction.

At sufficiently low frequencies the fields in the basic cell are identical with the static fields. The next step in the analysis is to derive a formula by which β may be calculated if $\boldsymbol{E}_0$ and $\boldsymbol{H}_0$ are known, which can be done by applying Poynting's theorem to the cell of Fig. 4. This theorem states

$$\int_S (\boldsymbol{E}_\wedge\boldsymbol{H})_\nu dS = -j\omega\int_V (\epsilon_0 E^2 + \mu_0 H^2)dV \quad . \quad (22)$$

where V is a volume enclosed by a surface S and ν is the direction of the outward normal to S at a point on S. Let V be the volume of the cell excluding that occupied by the conductor, so that S includes the six faces of the cell and the surface of the conductor. It is easily seen from the physical interpretation of the surface integral in terms of energy flow that the only contributions to the left of eqn. (22) arise from the faces $z = \pm c/2$. Hence

$$\int_S (\boldsymbol{E}_\wedge\boldsymbol{H})_\nu dS = \int_{-b/2}^{b/2}\int_{-a/2}^{a/2}\left\{[(\boldsymbol{E}_\wedge\boldsymbol{H})_z]_{z=c/2} - [(\boldsymbol{E}_\wedge\boldsymbol{H})_z]_{z=-c/2}\right\}dxdy \quad . \quad . \quad . \quad . \quad (23)$$

The notation $[\,.\;.\;.\,]_{z=c/2}$ means that everything within the bracket has to be evaluated for z equal to $c/2$. The integrand on the right may be considerably simplified by using eqns. (13) and (14) with the result

$$\int_S (\boldsymbol{E}_\wedge \boldsymbol{H})_v dS = [\exp(-2j\beta c) - 1]\left[\int_{-b/2}^{b/2}\int_{-a/2}^{a/2}(\boldsymbol{E}_\wedge \boldsymbol{H})_z dxdy\right]_{z=-c/2}$$

$$= -2j\beta c\left[\int_{-b/2}^{b/2}\int_{-a/2}^{a/2}(\boldsymbol{E}_\wedge \boldsymbol{H})_z dxdy\right]_{z=-c/2} \quad . \quad (24)$$

when c is very much less than the wavelength.

The approximation that $\boldsymbol{E}$ and $\boldsymbol{H}$ are equal to $\boldsymbol{E}_0$ and $\boldsymbol{H}_0$ may now be used in eqns. (22) and (24): the following equation then results:

$$2\beta_0 nc\left[\int_{-b/2}^{b/2}\int_{-a/2}^{a/2}(\boldsymbol{E}_{0\wedge}\boldsymbol{H}_0)_z dxdy\right]_{z=-c/2} = \omega\int(\epsilon_0 E_0^2 + \mu_0 H_0^2)dV \quad . \; . \; . \; . \; . \quad (25)$$

From eqns. (18) and (19),

$$(\boldsymbol{E}_{0\wedge}\boldsymbol{H}_0)_z = \frac{\partial\phi}{\partial x}\frac{\partial\psi}{\partial y} - \frac{\partial\phi}{\partial y}\frac{\partial\psi}{\partial x} \quad . \; . \; . \; . \; . \quad (26)$$

and
$$\int_{-b/2}^{b/2}\int_{-a/2}^{a/2}\left(\frac{\partial\phi}{\partial x}\frac{\partial\psi}{\partial y} - \frac{\partial\phi}{\partial y}\frac{\partial\psi}{\partial x}\right)dxdy = \int_{\psi(-b/2)}^{\psi(b/2)}\int_{\phi(-a/2)}^{\phi(a/2)} d\phi d\psi \quad . \quad (27)$$

since $\left(\frac{\partial\phi}{\partial x}\frac{\partial\psi}{\partial y} - \frac{\partial\phi}{\partial y}\frac{\partial\psi}{\partial x}\right)$ is the Jacobian function of ϕ and ψ with respect to x and y. Also ϕ is constant when $x = \pm a/2$ and ψ when $y = \pm b/2$;

$\phi(a/2) - \phi(-a/2) = \Phi$, the potential difference between the conductors, which corresponds to the electrostatic field, $\boldsymbol{E}_0$; and

$\psi(b/2) - \psi(-b/2) = I$, the current which produces the magnetic field $\boldsymbol{H}_0$, when flowing across a width b of the conductors. From the above results it follows that the left-hand side of eqn. (25) reduces to $2\beta_0 nc\Phi I$.

The right-hand side of eqn. (25) can be simplified by noting that $\epsilon_0\int_V E_0^2 dV$ and $\mu_0\int_V H_0^2 dV$ are respectively twice the electrostatic and magnetostatic energy stored in the cell. The first integral therefore equals $C\Phi^2$ where C is the capacitance of the cell measured between the conductors and is given by

$$C = \epsilon_0\epsilon_r bc/a \quad . \; . \; . \; . \; . \; . \; . \quad (28)$$

if ϵ_r is the relative permittivity of the artificial dielectric. Similarly, the second integral equals LI^2 where L, the inductance of the cell, is given by

$$L = \mu_0\mu_r ac/b \quad . \; . \; . \; . \; . \; . \; . \quad (29)$$

μ_r being the relative permeability of the artificial dielectric.

Eqn. (25) may now be written

$$2\beta_0 n\Phi I = \omega(\epsilon_0\epsilon_r\Phi^2 b/a + \mu_0\mu_r I^2 a/b) \quad . \; . \; . \quad (30)$$

It now remains to find a relation between Φ and I and this may be done by using Poynting's theorem in its complex form. This states

$$\int_S (\boldsymbol{E}_\wedge \boldsymbol{H}^*)_v dS = j\omega\int_V (\epsilon_0 \boldsymbol{E}.\boldsymbol{E}^* - \mu_0 \boldsymbol{H}.\boldsymbol{H}^*)dV \quad . \quad (31)$$

the asterisk denoting the complex conjugate. Once again the only contributions to the surface integral arise from the cell faces at $z = \pm c/2$ but this time these cancel since

$$[\boldsymbol{E}_\wedge \boldsymbol{H}^*]_{z=c/2} = [\boldsymbol{E}_\wedge \boldsymbol{H}^*]_{z=-c/2} \quad . \; . \; . \quad (32)$$

from eqns. (13) and (14). Hence,

$$\epsilon_0\int_V E_0^2 dV = \mu_0\int_V H_0^2 dV \quad . \; . \; . \; . \quad (33)$$

since $\boldsymbol{E}_0$ and $\boldsymbol{H}_0$, the first approximations to $\boldsymbol{E}$ and $\boldsymbol{H}$, are both purely real. This equation becomes

$$\epsilon_0\epsilon_r\Phi^2 b/a = \mu_0\mu_r I^2 a/b \quad . \; . \; . \; . \quad (34)$$

by the energy arguments used above. Φ and I may now be eliminated from eqn. (30) which becomes on substituting $\omega(\epsilon_0\mu_0)^{\frac{1}{2}}$ for β_0

$$n = (\epsilon_r\mu_r)^{\frac{1}{2}} \quad . \; . \; . \; . \; . \; . \; . \quad (35)$$

This is the result which it was desired to prove.

The above proof justifies the representation of the artificial dielectric by a transmission line for which the inductance and capacity per unit length are equal to the values, derived from static fields, per cell of the artificial dielectric. The actual result relating to the refractive index may be very simply derived if this representation is assumed, but in view of its importance it has been thought desirable to develop the proof direct from Maxwell's equations.

621.372.2:621.315.61

Paper No. 1700 R
Jan. 1955

SOME EXPERIMENTS ON ARTIFICIAL DIELECTRICS AT CENTIMETRE WAVELENGTHS

By M. M. Z. EL-KHARADLY, B.Sc., Ph.D., Graduate.

(*The paper was first received* 28*th January, and in revised form* 11*th May,* 1954.)

SUMMARY

A parallel-plate transmission line has been developed for wavelengths between 8 and 11 cm and has been used for an examination of the properties of artificial dielectrics under conditions which approximate closely to those existing in free space. An experimental method by which the effective electrical length of dielectric samples can be determined has been shown to give reliable results. Particular attention has been paid to the occurrence of dispersion and a distinction is drawn between two forms of this. The first arises from the resonance of the elements from which the dielectric is constructed and is called "dipolar dispersion" because of the similarity to the corresponding behaviour of solid dielectrics. The second depends primarily on the spacings between the conducting elements and is referred to as "cavity dispersion." Experimental results are given for both types of dispersion.

(1) INTRODUCTION

The principal uses of artificial dielectrics, i.e. arrays of conducting elements which affect electromagnetic waves in a way similar to a solid dielectric, arise in the centimetric region and a convenient method of making measurements on them at such wavelengths is obviously desirable. Hitherto the choice has been limited to making the measurements inside a waveguide or in free space and in either case there are considerable difficulties. Waveguide measurements give accurate results only if the artificial dielectric is effectively homogeneous within the guide,[1] and the spacings and sizes of the conducting elements are therefore restricted to much less than the wavelength being used.* The accuracy of free-space measurements depends on the care with which stray reflections are eliminated and a complicated arrangement of apparatus is invariably required. Further, the provision of sufficiently large samples of the artificial dielectric is both costly and time-consuming, especially when a wide range of element sizes and spacings is to be covered.

A modified method, which is a combination of the free-space and waveguide methods, has been developed and may be justified with the help of Fig. 1 illustrating a tetragonal array in which the elements are taken to be spheres: a plane electromagnetic wave polarized as shown is incident normally on the interface between the array and free space. If perfectly conducting plates of negligible thickness are inserted at right angles to the electric field in the planes mid-way between the layers of spheres, the field distributions are in no way altered. It is therefore possible to isolate the region between one pair of conducting plates and to make all the measurements within this region. It is well known that the only type of wave which can propagate freely between parallel plates, whose distance apart is less than half the free-space wavelength, is the TEM mode.[2] This is identical in all essential respects to a plane wave propagating in free space. It is desirable, to avoid complications from the possibility that more than one mode might propagate, to restrict the spacing between the plates to less than half the free-space wavelength: this is adequate for an investigation of the artificial dielectrics of immediate interest.

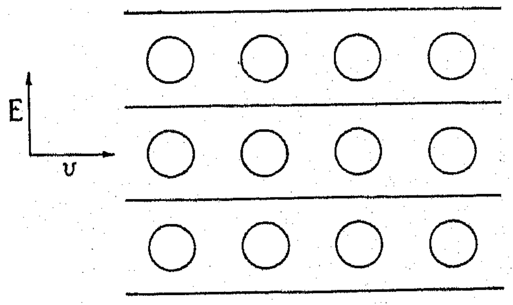


Fig. 1.—Tetragonal array of spheres.

The system outlined above requires what is effectively a parallel-plate transmission line with plates of infinite extent. In a practical version the width of the line must be made finite: it may be noted that a similar restriction is inevitable when free-space measurements are made, the cross-section of the plane wave being finite instead of infinite as theoretically required. From a study of the results obtained by using plane waves, it was decided that the minimum width for the transmission line should be five wavelengths. Since the properties of the artificial dielectrics depend on the ratio of the dimensions used to the wavelength, a choice of wavelength was available. Long wavelengths involve a bulky transmission line while short wavelengths require the provision of very small elements in the dielectrics. The range 8–11 cm was selected as a reasonable compromise between these two extremes and had the further advantage that convenient oscillators were readily available. The dimensions of the transmission line were accordingly fixed at a spacing of 3 cm, this ensuring that only the TEM mode propagates even at the shortest wavelength used, and a width of approximately 60 cm. The spacing between the plates can easily be changed to any value in the range 2·5–4·0 cm.

(2) DETAILS OF APPARATUS

The principal components of the apparatus are shown in Fig. 2. The transmission line is constructed from 3/8 in sheets of Bakelite, lined on the inner surface with tin foil. Each sheet is 60 cm square and successive sections are coupled together by flanges formed from 1 in angle pieces fixed around the edges. Provision is made for a total length of line of 465 cm. The spacing between the plates is maintained by wooden blocks along the edges of the line: these blocks are tapered towards the inside of the line and have little effect on the field distribution within, since the field strengths at the edges of the line are relatively low. A further useful property of the blocks is that they form a screen, preventing any external radiation from disturbing the fields within the line.

One section of the line is an accurately machined 3/8 in-thick

* In the special case of conductors of cylindrical cross-section, with their axes aligned perpendicular to the directions of the electric field and propagation, there is no restriction on the use of waveguides. This is referred to in Section 4.3.1.

Written contributions on papers published without being read at meetings are invited for consideration with a view to publication.

Dr. El-Kharadly, who was formerly in the Electrical Engineering Department of Imperial College of Science and Technology, University of London, is now in the Electrical Engineering Department of Ein-Shems (Ibrahim) University, Cairo, Egypt.

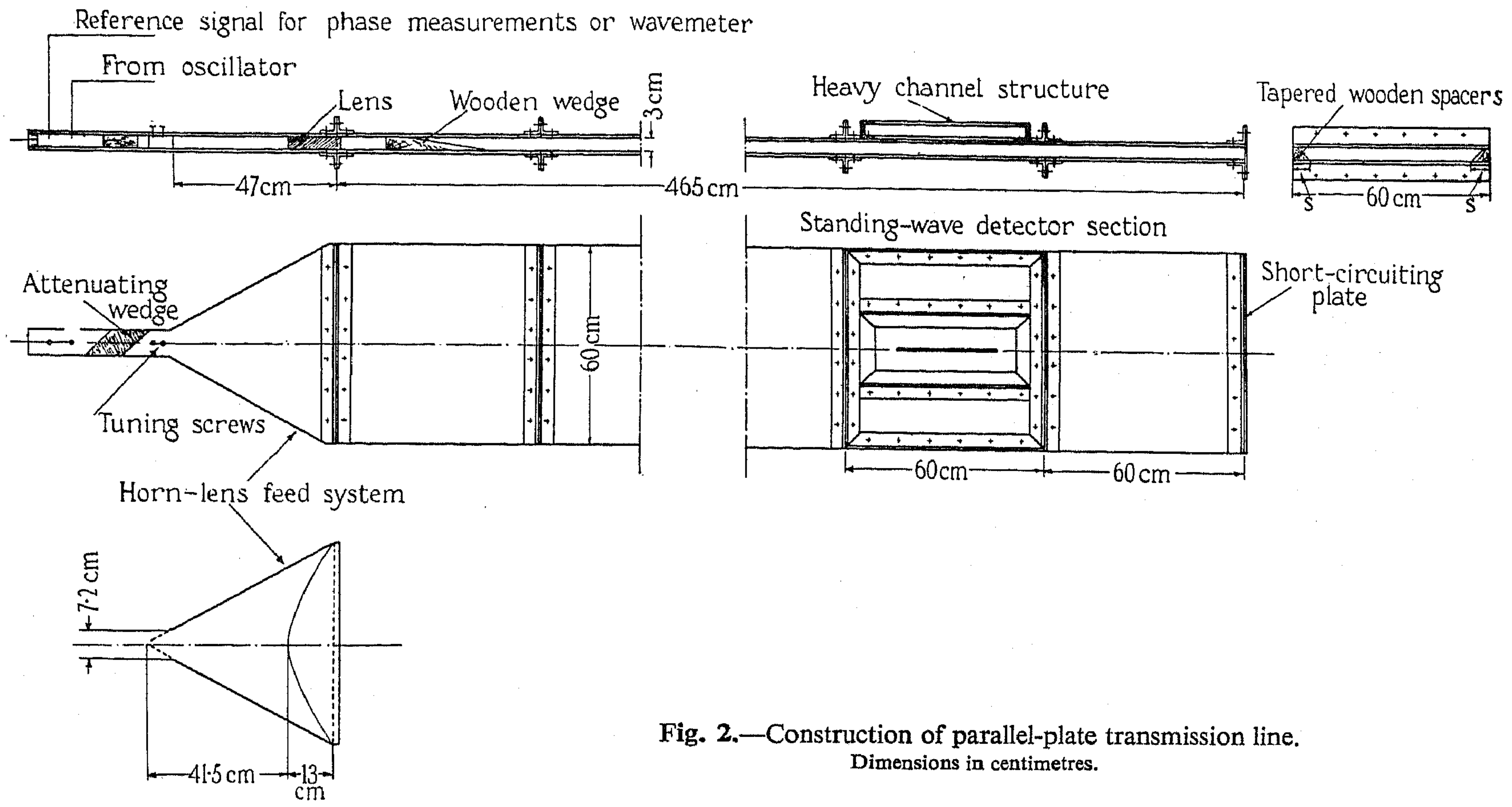


Fig. 2.—Construction of parallel-plate transmission line.
Dimensions in centimetres.

brass plate slotted for 30cm along its centre line, and enables a probe to be inserted into the transmission line. This plate is strengthened by rectangular-section, steel channel and flanges are provided so that the slot may be positioned parallel or perpendicular to the edges of the line. The probe and crystal detector, similar to those used in a waveguide standing-wave detector, are mounted on a soft-iron block, $3\frac{1}{2} \times 3\frac{1}{2} \times \frac{3}{4}$in, accurately machined to ensure constant probe-penetration. The position of the probe can be read from a scale and vernier to an accuracy of 0·1mm. For phase measurements a probe coupled direct to a coaxial cable can be used.

The required field distribution in the line is established with the help of a lens-corrected horn as shown. The lens is made from polystyrene and the curved surface is calculated to produce a plane-phase surface in the aperture. The signal from the oscillator is introduced by a probe inserted in the waveguide coupled to the horn. An attenuating pad is placed in this guide to decouple the oscillator, and further decoupling, to prevent the possibility of resonances in the transmission line, is provided in the form of a wooden attenuating strip placed in front of the lens aperture.

All measurements are made with the transmission line terminated in either a matched load or a short-circuit. The former consists of a tapered wooden wedge extending right across the line, and when it is used, a standing-wave ratio of better than 1·03 is obtained at any wavelength in the range 8–11 cm. The short-circuit is formed by bolting a flat brass plate of thickness 1/8 in to the flanges of the end sections. In view of the width of the line it has not been considered feasible to make a movable short-circuit.

(3) PRELIMINARY MEASUREMENTS

(3.1) Examination of the Field Pattern within the Line

The way in which the amplitude of the electric field varies along the direction of propagation was first investigated with the line terminated in the matched load, and a marked variation was found to occur near the lens. This corresponds to Fresnel interference near a radiating aperture and the variation became much less as the distance from the lens was increased. It was found experimentally that the amplitude was sensibly constant, as shown by the plots in Fig. 3, when the length of the line was more than 400cm. The phase under the same conditions varied linearly with a slope appropriate to the wavelength used.

Transverse-amplitude and phase patterns, again taken under matched-load conditions at about 400cm from the lens aperture,

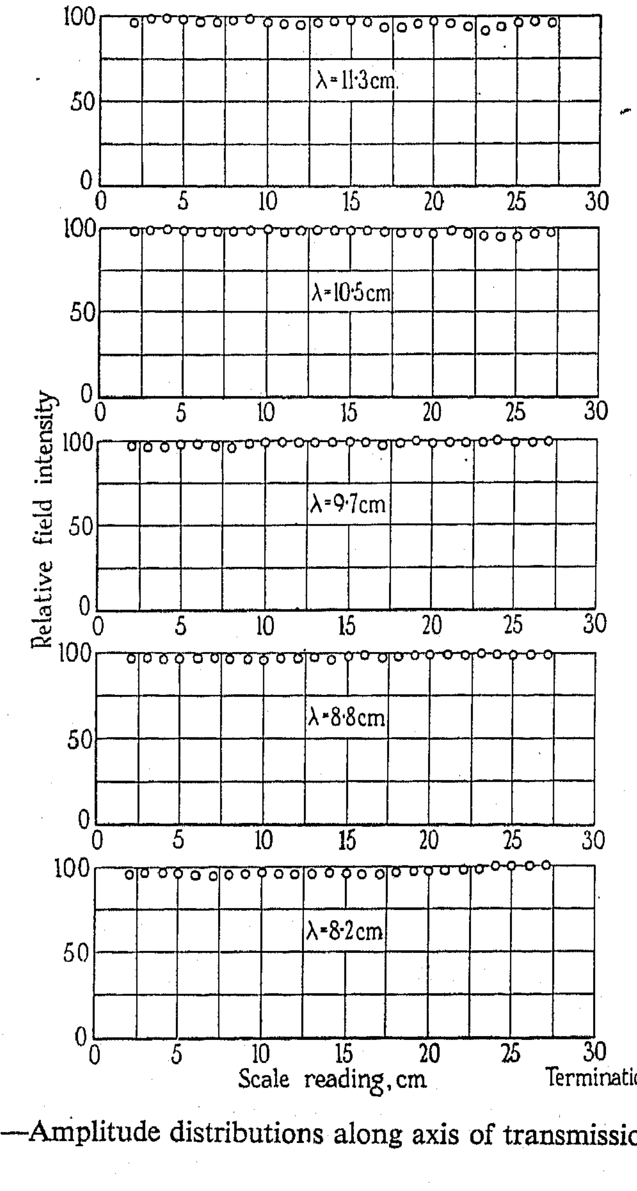


Fig. 3.—Amplitude distributions along axis of transmission line.

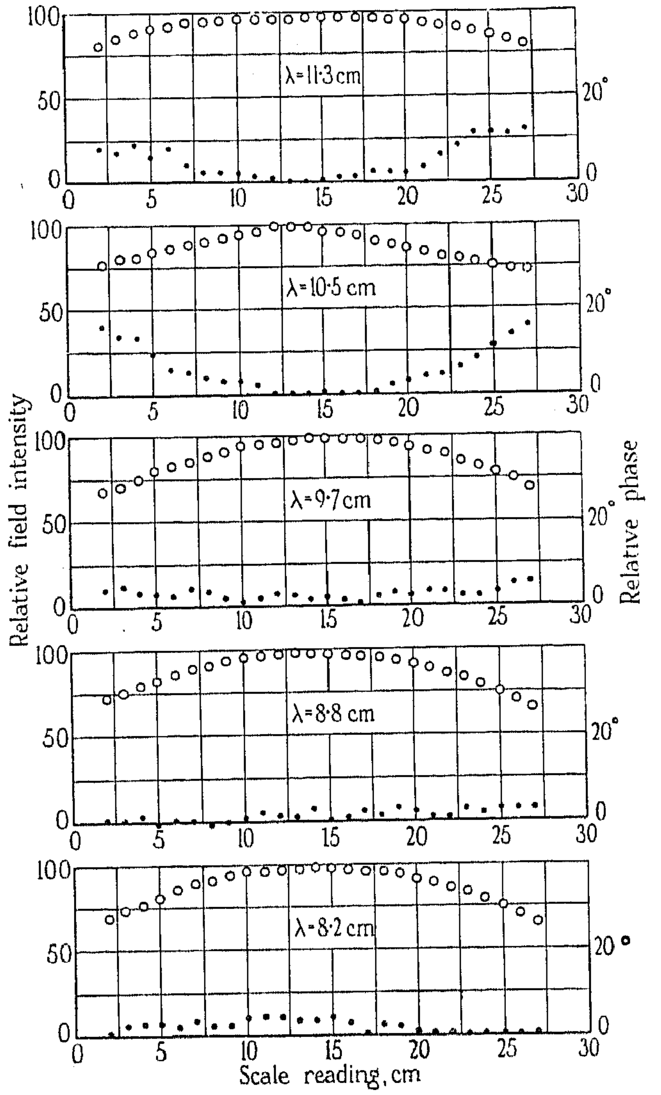


Fig. 4.—Amplitude and phase distributions across transmission line.
○○○ Amplitudes.
●●● Phases.

are shown in Fig. 4. While the measurements for these were being taken, the parts of the slot, not covered by the standing-wave-detector block, were closed by cylindrical rods to eliminate the radiation which would otherwise occur. The taper of the amplitude patterns across the line over the 30cm range of the slot, is not considered sufficiently serious to invalidate the assumption that near the centre of the line the field is basically that of a TEM mode. It was verified by additional measurements carried out with a probe moving across the open end of the line, that the field strength became very small at the edges of the line (see Fig. 5). The pattern given in this figure suggests that the field approximates to the H_{01} mode for a rectangular waveguide of width 50cm: calculation then shows that the wavelength measured within the line should not differ from the free-space wavelength by more than 0·5%. This was confirmed experimentally from the distance between successive minima when the line was terminated by a short-circuit.

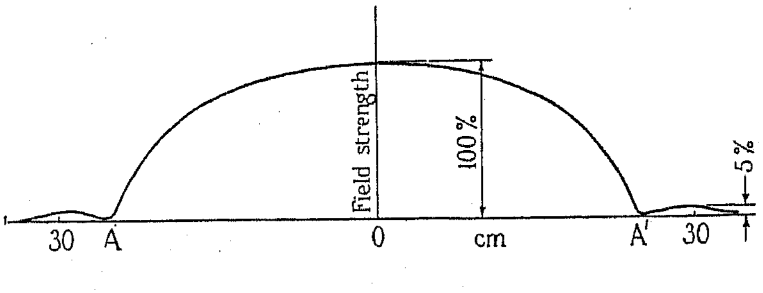


Fig. 5.—Complete amplitude distribution across transmission line.

The phase measurements were all carried out by the standard method of mixing the signal from an r.f. probe with a reference signal and are estimated to be correct to within ±2°.

(3.2) Dielectric Measurements

Since the primary object of the transmission line is to carry out an investigation of artificial dielectrics, it was thought advisable to confirm that it gave results for solid dielectrics in agreement with those obtained by other methods. The relative permittivities of polystyrene and expanded polystyrene were obtained by an application of the Roberts–von Hippel method: slabs of the material of width 50cm were placed across the transmission line in contact with the short-circuit and the change in the position of the field minimum was observed. The relative permittivity was calculated in the usual way and measurements were made at five wavelengths between 8 and 11cm. For expanded polystyrene, values between 1·036 and 1·038 were obtained for the relative permittivity, the corresponding result for a waveguide measurement being 1·035. The values for polystyrene ranged from 2·49 to 2·59, but in the restricted range of wavelengths from 8·80 to 10·50, all the values lay between 2·49 and 2·52. The corresponding value from a waveguide measurement was 2·54. It thus appears that the accuracy obtainable for relative permittivity is not worse than 2% over the major part of the wavelength range used.

A further check was obtained by making measurements on artificial dielectrics for which it was expected that neither dispersion nor the interface effect discussed in Section 4 would affect the results. Theoretical values for comparison are thus available from measurements carried out on similar dielectrics at low frequencies.[3] The method of measurement was extended to enable both the refractive index and the normalized wave impedance of the medium to be obtained: this involved determining the minima position when the medium was backed by either the short-circuit or an effective open-circuit, i.e. a short-circuit placed at a distance of a quarter of the free-space wavelength, λ, from the medium. The theory of the method is discussed in Section 9. The experimental values of the refractive index n, and the normalized wave impedance Z, are shown in Fig. 6, for a cubical lattice of $\frac{1}{4}$in diameter spheres, the lattice spacing being 1cm, and in Fig. 7 for a cubical lattice, spacing 1·5cm, of $\frac{1}{2}$in diameter discs made from copper foil of thickness 0·002in. In Fig. 6 the theoretical values of n and Z, obtained from the low-frequency measurements, are shown by solid lines. The agreement for the refractive index is within 1%, except that there is a slight increase for the shorter wavelength in the case of the disc array. This indicates that dispersion is beginning to be important. The values of Z which depart most from the theoretical curve are those for sample lengths equal to a multiple of $\lambda_0/4n$. It is shown in Section 9 that the

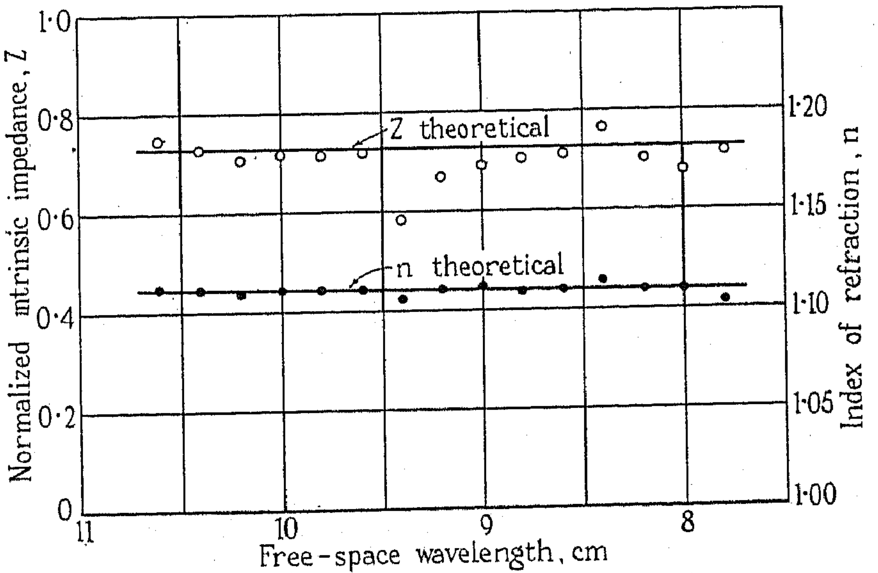


Fig. 6.—Refractive index and normalized impedance for a cubical lattice of spheres.

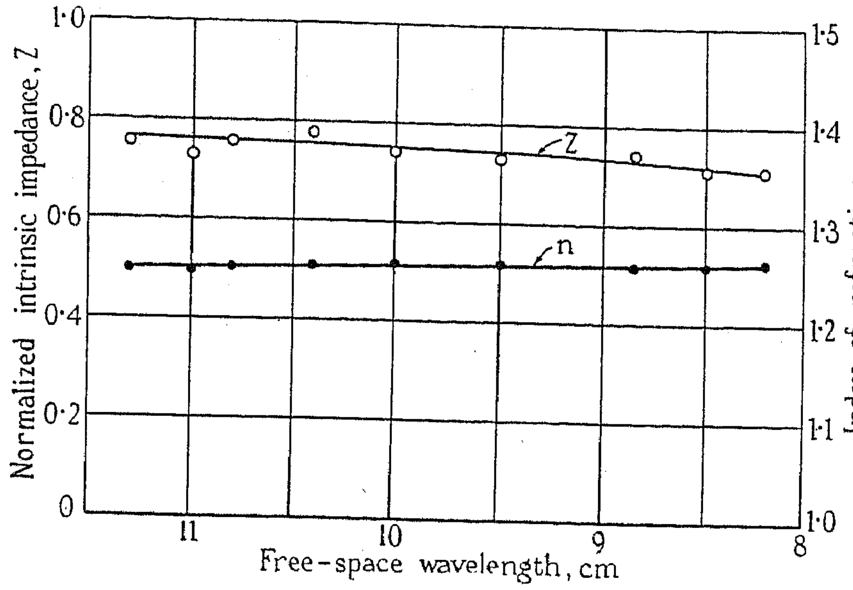


Fig. 7.—Refractive index and normalized impedance for a cubical lattice of discs.

accuracy of the method then deteriorates. The sample lengths used for the measurements were 15 cm and 21 cm for the sphere and disc arrays respectively.

(4) THE DETERMINATION OF THE EFFECTIVE INTERFACE BETWEEN FREE SPACE AND AN ARTIFICIAL DIELECTRIC

(4.1) General Discussion

There is no clear-cut boundary between free space and an artificial dielectric, and it is quite possible for fairly large errors in the values of refractive index and normalized impedance obtained by the method described in Section 9 to arise because of uncertainty as to the actual position of the interface. In the simplest case, the interface can be shown to lie a distance $b/2$ from the plane through the centres of the first layer of obstacles, where b is the spacing between successive rows of elements. This has been assumed in evaluating the results of the preliminary experiments described in the previous Section. For media having closely spaced obstacles, relatively complicated variations in the field distribution occur near the interface, and allowance for this must be made in the interpretation of measurements of the properties of such media. A qualitative examination of such interface effects is given in the introductory paper by Brown and Jackson.[4] The simplest procedure is based on the assumption that the interface plane is shifted a distance Δ from the position discussed above towards the first layer of obstacles. An experimental method by which this shift may be determined is described in Section 4.2.

(4.2) Experimental Determination of Effective Length

The most obvious effect of the shift in the position of the interface is a change in the electrical length of the sample by an amount 2Δ. In the interpretation of measurements, however, the shift also causes changes of magnitude Δ in the distances of the field minimum and of the short-circuit from the interfaces at the front and rear of the sample, respectively. The calculation of the refractive index and of the normalized impedance must accordingly be modified to allow for each of these effects.

The reactive field associated with the interface extends beyond the region occupied by the medium and is liable to be altered if a short-circuit is placed immediately behind the sample. Accordingly the short-circuit and open-circuit measurements were carried out with the short-circuiting end-plate distant $\lambda_0/2$ and $3\lambda_0/4$ respectively from the effective interface at the rear of the sample. These distances are sufficient to ensure that the interface reactive-field is negligible at the short-circuiting end-plate.

The shift Δ is of course not known and has to be measured. This may be done by carrying out measurements, as described in Section 9, for each of two samples of different lengths and for a series of assumed values of Δ. The normalized impedance is sensitive to changes in Δ and if the values obtained from the two samples are plotted against Δ, curves of the type shown in Fig. 8 result. These curves intersect at some point and this gives the true values of Δ and Z.

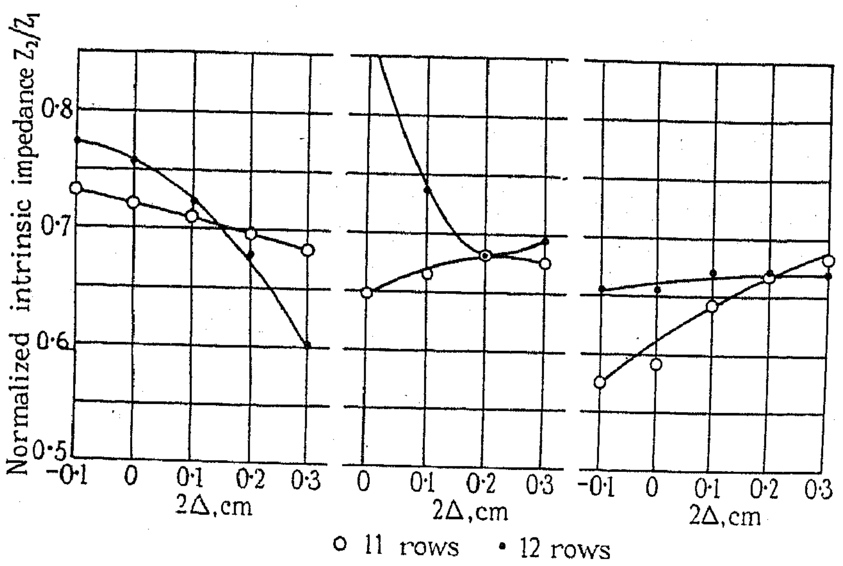


Fig. 8.—Curves obtained in determination of electrical length of medium.

This method has been applied to an artificial dielectric composed of 1/16 in-diameter rods of length 1 in, arranged in a lattice with ¼ in spacings in the direction of propagation and 1 in spacings in the transverse direction. The two samples used had 11 and 12 layers of rods. Typical curves showing the variation of the calculated impedance with the assumed value of Δ are shown in Fig. 8, and the shift Δ is plotted against wavelength in Fig. 9.

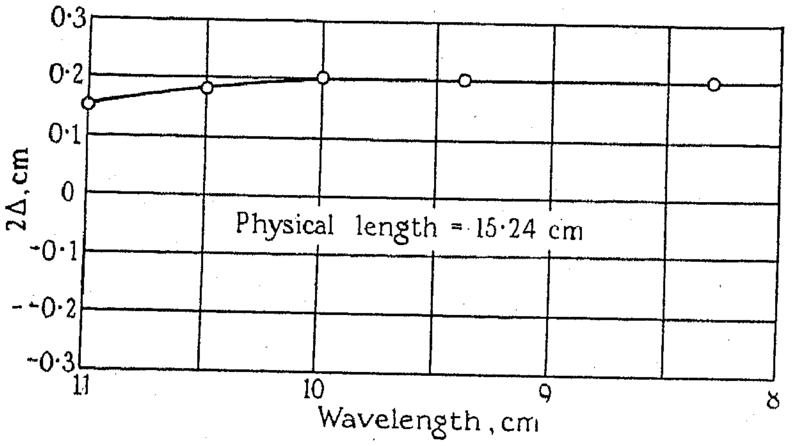


Fig. 9.—Dependence of change in length on wavelength.

Despite the fact that the method described in this Section is based on a very simple assumption, which only approximates to the complete interface effect, it does lead to a considerable improvement in the consistency of the results obtained for the normalized impedance. This is illustrated by Fig. 10, in which the points marked with large circles are those obtained by the present method while those marked by crosses result from assuming Δ to be zero. The former set shows a much smoother variation with wavelength.

(5) DISPERSIVE PROPERTIES OF ARTIFICIAL DIELECTRICS

(5.1) General Considerations

It has so far been generally accepted that the effective permittivity and permeability of artificial dielectrics are independent

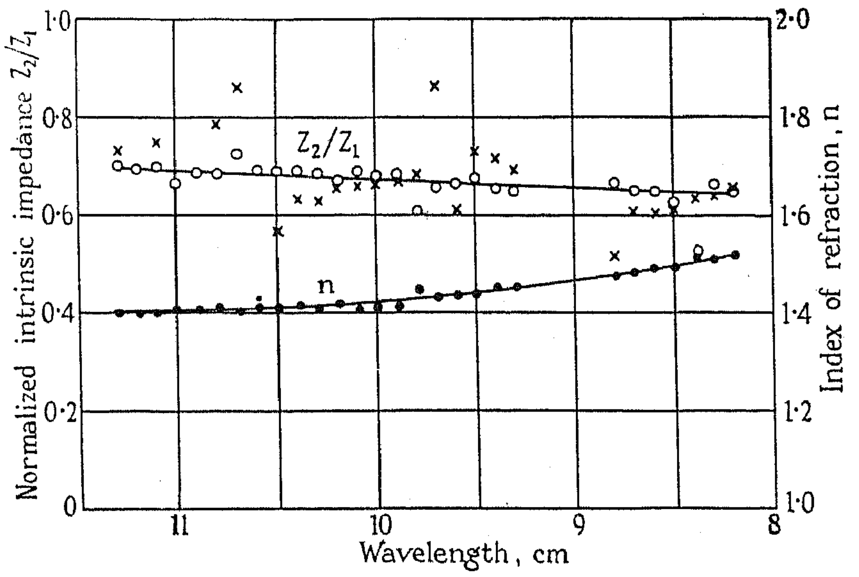


Fig. 10.—Results obtained if electrical length of specimen is determined.

of the wavelength used so long as it is much greater than all the dimensions and spacings of the elements forming the dielectric. A commonly accepted criterion is that the wavelength should exceed ten times the largest dimension involved: it may be noted that on this basis dispersion may be expected with the disc array examined in Section 3.2. The methods by which the low-frequency values of permittivity and permeability may be obtained have already been discussed at length.[3]

A broad distinction may be drawn between the cases when the dispersion arises because of the frequency-dependent properties of the elements of the array and when it arises because the spacings become comparable with the wavelength of operation. The former is the more akin to what happens in a solid dielectric and may be termed "dipolar dispersion." In the latter, the region between adjacent elements behaves similarly to a cavity resonator of low Q-factor and the description "cavity dispersion" is appropriate. There is no clear-cut distinction between the two types of dispersion and the dispersive properties of many artificial dielectrics arise from both causes.

(5.2) Dipolar Dispersion

The possibility of dipolar dispersion was mentioned by Kock in his original paper on metallic delay dielectrics.[5] It occurs when the wavelength approaches the resonant wavelength of the basic element from which the dielectric is formed. For example, a perfectly conducting sphere of diameter d has a resonant wavelength equal to $2\pi d/\sqrt{(3)}$,[6] while for long thin spheroids the resonant wavelength is approximately twice the major axis.[7] A first approximation to the dependence of the relative permittivity, ϵ_r, on the wavelength of operation λ_0, may be found from the analogy to solid dielectrics. The simplest low-frequency formula is[3]

$$\epsilon_r = 1 + aN/\epsilon_0 \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

where a is the polarizability of each element,
N is the number of elements per unit volume
and ϵ_0 is the free-space permittivity.

Near the resonant wavelength λ_r this is modified to

$$\epsilon_r = 1 + \frac{aN/\epsilon_0}{1 - (\lambda_r/\lambda_0)^2} \quad . \quad . \quad . \quad . \quad . \quad (2)$$

For very long wavelengths, eqns. (1) and (2) are identical. As λ_0 decreases towards λ_r, ϵ_r increases and becomes infinite when λ_0 and λ_r are equal. If λ_0 is less than λ_r, ϵ_r may become negative: this implies that propagation through the dielectric is no longer possible.

At very long wavelengths, eqn. (1) may be modified to allow for interaction between elements: a corresponding modification to eqn. (2) gives[8]

$$\epsilon_r = 1 + \frac{aN/\epsilon_0}{1 - AaN/\epsilon_0 - (\lambda_r/\lambda_0)^2} \quad . \quad . \quad . \quad (3)$$

where A is an interaction factor,[3] which may be taken as having its low-frequency value.

The principal change carried by eqn. (3) is that ϵ_r becomes infinite at a wavelength λ_r' given by

$$\lambda_r' = \lambda_r/[1 - AaN/\epsilon_0]^{\frac{1}{2}} \quad . \quad . \quad . \quad . \quad . \quad (4)$$

Since the interaction factor A depends on the spacings between the elements of the array, so also does the wavelength, λ_r', at which ϵ_r tends to infinity. In practice, infinite values of ϵ_r do not arise because of the finite conductivity of the metals from which the elements are made.

(5.3) Cavity Resonance

The dependence on wavelength of the relative permittivity, or more conveniently the refractive index, has already been examined for the artificial dielectrics whose elements have negligible thickness in the direction of propagation. Examples of such dielectrics are arrays of strips or discs and they can be represented, under certain restrictions on the lattice spacings, by a transmission line loaded at regular intervals by shunt susceptances.[9,10,11] The refractive index, n, may then be calculated from the equation[10]

$$\cos \beta b = \cos \beta_0 b - \tfrac{1}{2} B \sin \beta_0 b \quad . \quad . \quad . \quad (5)$$

where β_0 (equal to $2\pi/\lambda_0$) and β (equal to $\beta_0 n$) are the phase constants for propagation in free space and within the dielectric respectively, B is the normalized susceptance of each plane of elements, and b is the spacing between successive planes of elements. Eqn. (5) is valid if the susceptances are independent, i.e. if there is no reactive coupling between adjacent planes of elements. It has been found that this is so if the spacing b exceeds the spacings between the elements in the plane at right angles to the direction of propagation.[10, 11]

It can be shown from eqn. (5) that there are a succession of "stop" bands in which propagation is not possible.[12] An important practical case is that when B is capacitive, and it can then be shown that the refractive index n increases steadily as λ_0 decreases from very large values to a cut-off wavelength λ_c, which marks the beginning of the first stop band.

The normalized wave impedance, Z, of the artificial dielectric is given by the equation

$$Z = \frac{\tan(\pi b/\lambda_0)}{\tan(\pi b n/\lambda_0)} \quad . \quad . \quad . \quad . \quad . \quad (6)$$

which is valid under the same conditions as eqn. (5). If b is very much less than λ_0, Z reduces to $1/n$ in agreement with the value for a solid dielectric of refractive index n.

Eqn. (5) may be generalized to allow for any reactive interaction between adjacent planes of elements but becomes considerably more complicated.[9] It is also possible to carry out a similar analysis for elements of finite thickness in the direction of propagation by using more elaborate equivalent circuits as discussed in the "Waveguide Handbook."[13] If the elements used for the construction of the dielectric have a finite conductivity, it is possible to have an exact equivalent circuit only if resistive elements are included. Such resistances normally have an appreciable effect on the results only for wavelengths very near to the cut-off value.

One interesting application of eqn. (5) arises when B represents a series resonant circuit. This will happen if the elements themselves become resonant, i.e. if there is dipolar resonance. The susceptance of a series resonant circuit may be written

$$B = K/\lambda_0[1 - (\lambda_r/\lambda_0)^2] \quad . \; . \; . \; . \; . \quad (7)$$

where λ_r is the wavelength for which the circuit is resonant and K is a constant depending on the L/C ratio for the circuit. Suppose now that b is much less than λ_0 or λ_r and approximate accordingly in eqn. (5) which becomes[10]

$$n^2 = \frac{\beta^2}{\beta_0^2} = 1 + \frac{\lambda_0 B}{2\pi b} \quad . \; . \; . \; . \; . \quad (8)$$

Substitution for B from eqn. (7) gives

$$n^2 = 1 + \frac{K}{2\pi b[1 - (\lambda_r/\lambda_0)^2]} \quad . \; . \; . \; . \quad (9)$$

which is identical so far as the dependence on wavelength is concerned to eqn. (2). It is thus seen that the cavity dispersion formula reduces to the dipolar formula when the element spacings are small compared with the wavelength. This emphasizes that the distinction between the two types of dispersion is by no means an absolute one.

(5.4) Experimental Results

(5.4.1) Strip Dielectric.

The simplest artificial dielectric for which eqn. (5) is applicable consists of a two-dimensional array of thin conducting strips directed normally to both the directions of the electric field and of propagation. It is unnecessary in this case to use the transmission line since there is no essential change caused by placing the dielectric in a waveguide, provided that the guide wavelength is used in place of the free-space wavelength. The particular specimen investigated consisted of a succession of ten strips having dimensions as shown in Fig. 11 and spaced apart at regular intervals of 3·96 cm. The strips were made from tin foil of thickness 0·002 in, and were supported by blocks of expanded polystyrene which completely filled the waveguide. The method of measurement used was that described in Section 9 and the results obtained are plotted in Fig. 12 against the wavelength in the air-filled guide. The theoretical curves are calculated from eqns. (5) and (6) corrected to allow for the presence of Polyfoam: a value for the normalized susceptance of the strips is available in the "Waveguide Handbook."[13]

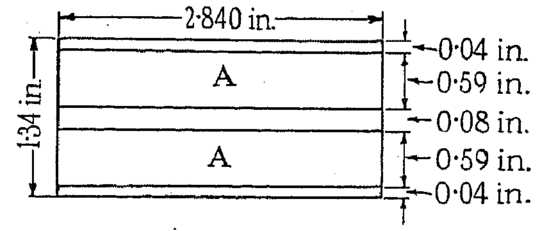


Fig. 11.—Dimensions of strips used in waveguide measurements.

Waveguide cross-section: 2·84 × 1·34in.

The agreement between the theoretical and experimental values plotted in Fig. 12 is on the whole good. The general trends of the theoretical curves are faithfully followed by the measured points and the maximum difference in numerical values is about 0·02 for the refractive index and 0·05 for the normalized impedance. The differences in the impedance values are largely caused by the effect discussed in Sections 3.2 and 9. For the shorter wavelengths, the measured refractive-index values are consistently 0·02 less than those calculated: this is probably due to slight errors in the dimensions and spacings of the strips.

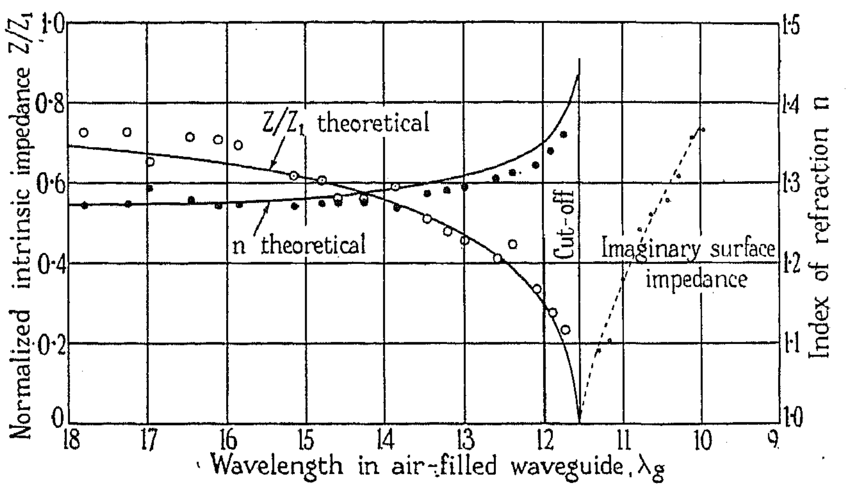


Fig. 12.—Refractive index and normalized impedance of strip dielectric.

For wavelengths less than the cut-off value, propagation within the strip dielectric is impossible, and both n and Z as deduced from eqns. (15) and (16) become imaginary. The attenuation constant can be deduced from the imaginary value of n but could not be measured with the equipment used. The bulk of the incident energy is reflected at the interface between the empty guide and the part filled with the dielectric: the imaginary value of Z is the effective reactance at this interface and can be readily deduced from the position of a field minimum in the empty guide. It may be seen from Fig. 12 that there is good agreement between the theoretical and measured values of Z except in the immediate vicinity of the cut-off wavelength. The conductivity of the strips has been ignored in the theoretical analysis and the agreement obtained between theory and experiment shows that this is justified except near cut-off. The conditions then are similar in some respects to those in a tuned circuit at resonance, and it is to be expected that the resistance of the strips will become important. It is difficult to make allowance for this theoretically, but it is very probable that the failure of the impedance to drop to zero at the cut-off wavelength is due to this resistance. Experimental confirmation of this was found from the measured values of the standing-wave ratio: at all wavelengths except those near cut-off the standing-wave ratio was too nearly zero for accurate measurement, but near cut-off it increased very rapidly to a maximum value of 0·4, showing that there was an appreciable absorption of power by the dielectric.

(5.4.2) Arrays of Rods.

The investigations so far carried out on the transmission line have been upon arrays of thin copper rods, directed parallel to the electric vector of the incident field, as illustrated in Fig. 13.

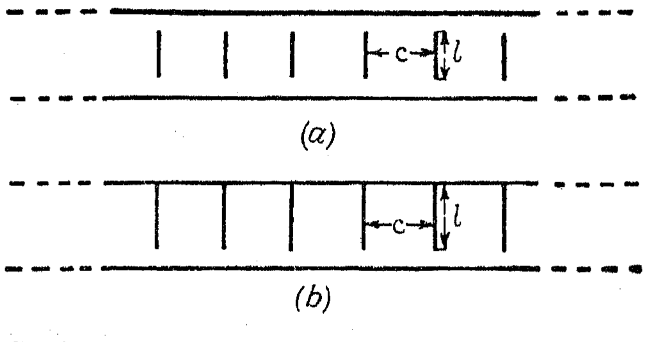
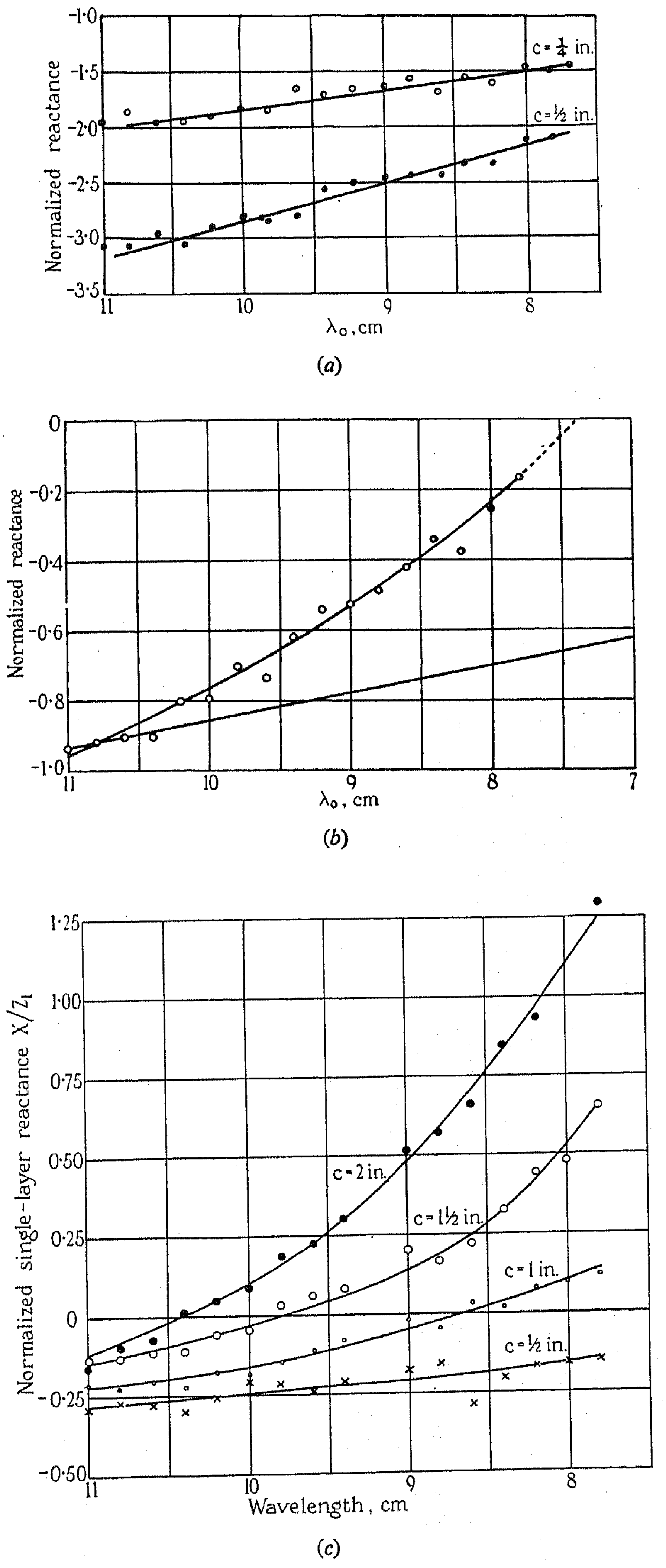


Fig. 13.—Section of transmission line through the axis of a single row of rods.

Two arrangements were used to permit a wide variation of the resonant wavelength of an isolated rod. In (*a*) the rods were supported by a slab of expanded polystyrene midway between



(*a*)

(*b*)

(*c*)

Fig. 14.—Normalized reactance of single rows of rods (diameter $\frac{1}{16}$ in):

	Arrangement of rods: see Fig. 13	Length cm	Resonant wave-length of isolated rod: cm	Spacing in
(*a*)	a	1·7	3·4	¼ and ½
(*b*)	b	2·0	8·0	1
(*c*)	b	2·54	10·2	½, 1, 1½ and 2

the plates of the transmission line, the resonant wavelength of an isolated rod being approximately twice its length. In (*b*) the rods had flat heads backed by tin foil to ensure good contact with one plate of the transmission line and were thus imaged in this plate. The resonant wavelength for an isolated rod of this type is approximately four times its length.

(5.4.3) Measurements on Single Rows of Rods.

No theoretical values for the reactance of a single row of rods are available, and preliminary measurements of this reactance were made. The diameter of the rods used was 1/16 in and it was considered that this was sufficiently small to permit the representation of a single row of rods by a shunt reactance. The row of rods, supported in a slab of expanded polystyrene, was placed a quarter-wavelength from the short-circuit terminating the line, and the distance, x, of a minimum in the electric-field distribution from the row of rods was measured. The normalized reactance was then calculated from the equation

$$X = -\tan(2\pi x/\lambda_0) \quad . \quad . \quad . \quad . \quad . \quad (10)$$

λ_0 being the wavelength in the transmission line. Measurements were carried out in this way on a number of rows of rods, and the normalized reactances are plotted in Fig. 14.

The resonant wavelength of an isolated rod in the arrangement used in Fig. 14(*a*) is much less than the actual wavelengths in the transmission line, and so the reactance is almost purely capacitive and therefore proportional to λ_0. In Fig. 14(*b*), the isolated rods resonate about 8·0 cm and the reactance shows a marked departure from a pure capacitance, which would behave as shown by the straight line in Fig. 14(*b*). The row of rods is resonant when its reactance becomes zero and it is seen that this happens at about 7·4 cm, which differs from the resonant wavelength (8·0 cm) for an isolated rod. This agrees with the theoretical predictions discussed in Section 5.2. The wide variation in the resonant wavelengths of the row as the spacing is changed is illustrated by Fig. 14(*c*).

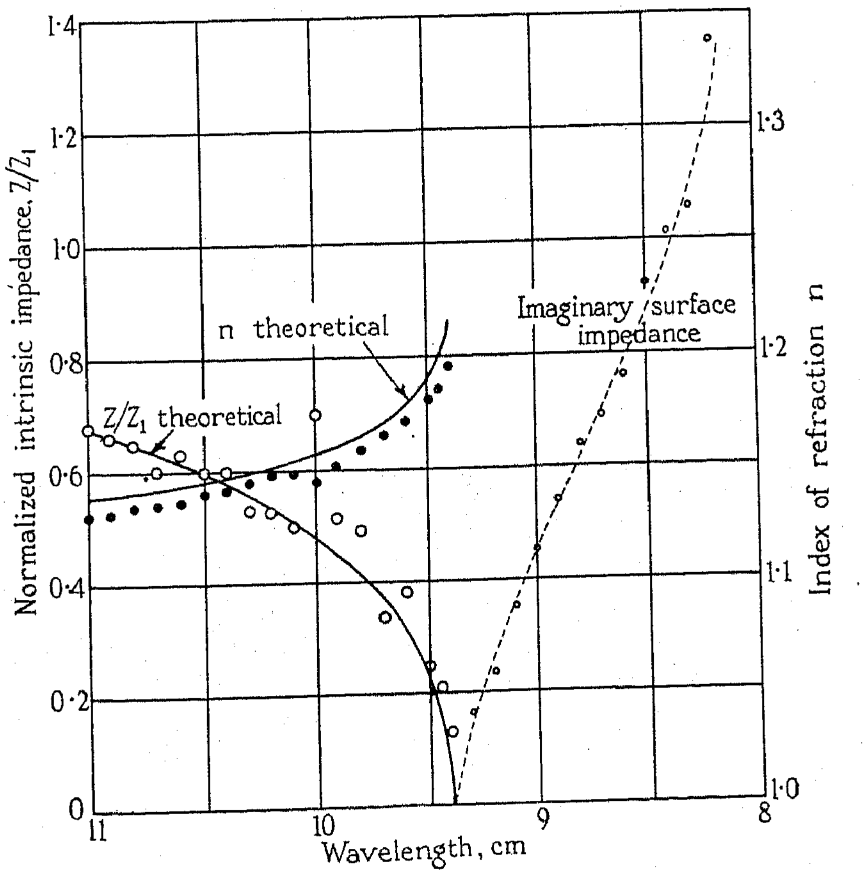
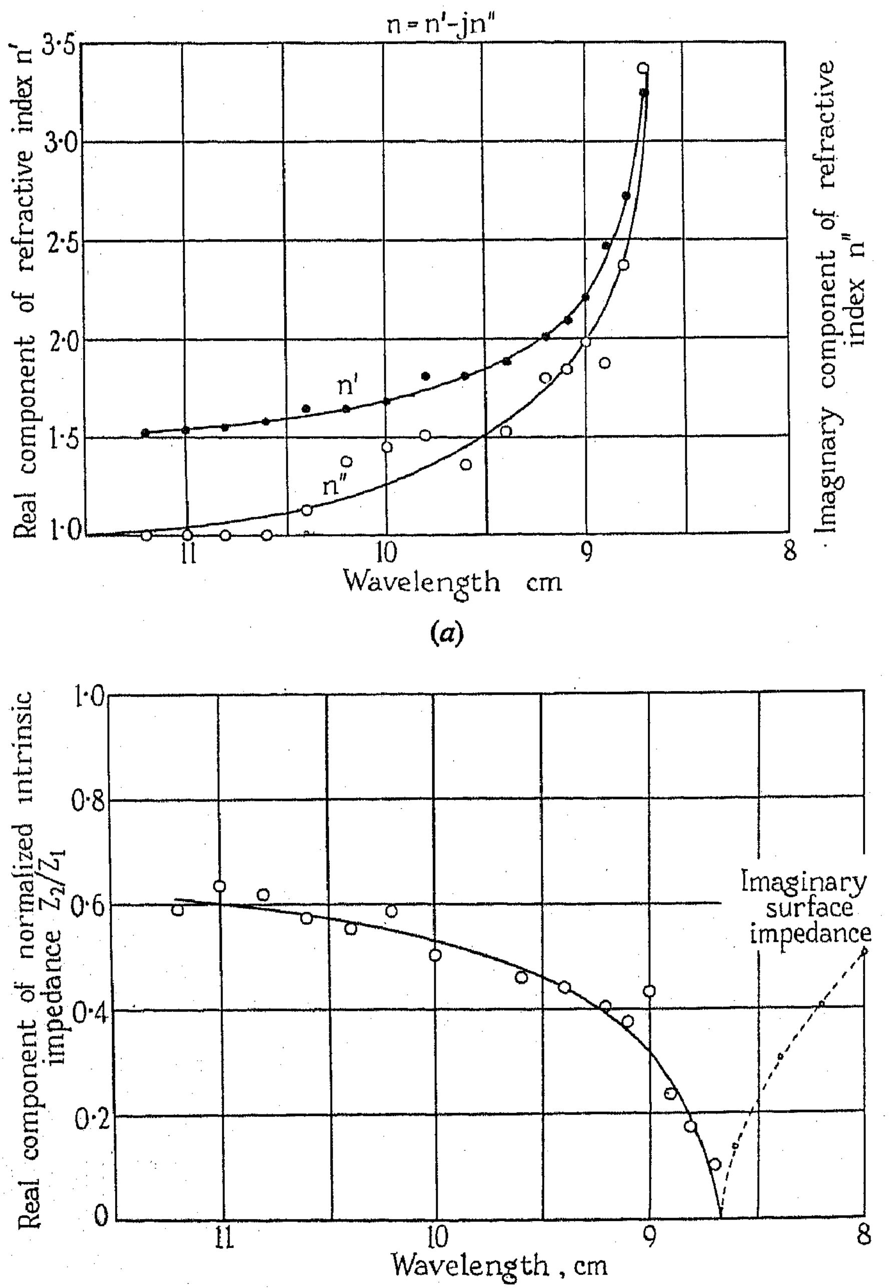


Fig. 15.—Refractive index and normalized impedance of dielectric formed from eight rows of rods as in Fig. 14(*a*) with spacing ½ in: spacing between rows 1½ in.

(5.4.4) Measurements on Arrays of Rods.

An artificial dielectric was assembled from eight rows of rods of the dimensions given for Fig. 14(*a*)—the spacing *c* being $\frac{1}{2}$in—and the spacing between successive rows being 1·5 in. The measured values of the refractive index and normalized impedance of this array are plotted in Fig. 15 together with theoretical curves calculated from the reactance values in Fig. 14(*a*). Since the short rods are almost purely capacitive, there is little difference between this case and the strip dielectric considered in Section 5.4.1.

The use of longer rods in the arrangement shown in Fig. 13(*b*) gives the possibility of dipolar dispersion. A dielectric was formed from twelve rows of the type considered in Fig. 14(*b*), the spacing between rows being $\frac{1}{4}$in. Since this spacing is only half that between the rods in a row, the effective electrical length of the specimen may differ from its physical length. The reasons for this and a method by which the electrical length may be measured have been discussed in Section 4. The measured values of the real and imaginary parts of the refractive index of this dielectric are plotted in Fig. 16(*a*): the imaginary part is proportional to the attenuation and increases rapidly as the resonant wavelength is approached, showing that the resistance of the rods becomes important near resonance. The normalized impedance, plotted in Fig. 16(*b*), shows the same general behaviour as for the dielectrics considered earlier.



(*a*)

(*b*)

Fig. 16.—Refractive index and normalized impedance of dielectric formed from twelve rows of rods as in Fig. 14(*b*): spacing between rows $\frac{1}{4}$in.

The resonant wavelength for this dielectric, as indicated by maximum refractive index and minimum impedance, is approximately 8·7 cm, which is considerably longer than the value shown in Fig. 14(*b*) for a single row of the rods. This difference in the resonant wavelengths can be explained in either of the following ways. It has already been established by the results in Fig. 14 that the resonant wavelengths of an isolated rod and a single row of rods differ because of the interaction field which is present in the second case but not in the first. Similarly, if more than one row are present, the interaction field is modified and the resonant wavelength is again changed. An alternative explanation is that cavity dispersion occurs as well as dipolar dispersion, and so the resonant wavelength must depend on the spacing between adjacent rows as predicted by the theory in Section 5.3. The observed result is, therefore, to be expected, whether the resonance is attributed to purely dipolar dispersion or to a mixture of dipolar and cavity dispersion.

(6) CONCLUSIONS

The possibility of using a parallel-plate transmission line to investigate the properties of artificial dielectrics has been demonstrated and values of the refractive index have been achieved with errors of less than 2%. A method of measuring the precise location of the effective interface between free space and an artificial dielectric has been developed and gives a considerable improvement in the consistency of the results for the wave impedance. The dependence of the refractive index and the wave impedance on the wavelength of operation has been measured and is shown to be in qualitative agreement with simple theories.

(7) ACKNOWLEDGMENTS

The work described is an extension to high frequencies of the investigation reported in Reference 3, and the author wishes to express his gratitude to Dr. Willis Jackson and Mr. John Brown for the guidance he received from them and for their assistance in the preparation of the paper.

(8) REFERENCES

(1) Corkum, R. W.: "Isotropic Artificial Dielectrics," *Proceedings of the Institute of Radio Engineers*, 1952, **40**, p. 574.
(2) Silver, S., et al.: "Microwave Antenna, Theory and Design" (McGraw-Hill Book Co., 1949.)
(3) El-Kharadly, M. M. Z., and Jackson, Willis: "The Properties of Artificial Dielectrics comprising Arrays of Conducting Elements," *Proceedings I.E.E.* (Paper No. 1472 R, July 1953), **100**, p. 199.
(4) Brown, J., and Jackson, Willis: "The Properties of Artificial Dielectrics at Centimetre Wavelengths" (see page 11).
(5) Kock, W. E.: "Metallic Delay Lenses," *Bell System Technical Journal*, 1948, **27**, p. 58.
(6) Debye, P.: "Der Lichtdruck auf Kugeln von beliebigen Material," *Annalen der Physik*, 1909, **30**, p. 57.
(7) Stratton, J. A., and Chu, L. J.: "Steady-State Solutions of Electromagnetic Field Problems," *Journal of Applied Physics*, 1941, **12**, p. 241.
(8) Joos, G.: "Theoretical Physics" (Blackie and Son, 1934), Book 4, Chapter 4.
(9) Cohn, S. B.: "Analysis of the Metal-Strip Delay Structure," *Journal of Applied Physics*, 1949, **20**, p. 257.
(10) Brown, J.: "The Design of Metallic Delay Dielectrics," *Proceedings I.E.E.*, 1950, **97**, Part III, p. 45.
(11) Cohn, S. B.: "Electrolytic-Tank Measurements for Microwave Delay Lens Media," *Journal of Applied Physics*, 1950, **21**, p. 674.

(12) Slater, J. C.: "Microwave Electronics" (Van Nostrand, 1950), Chapter 8.
(13) Marcuvitz, N.: "Waveguide Handbook" (McGraw-Hill Book Co., 1951).
(14) Roberts, S., and Von Hippel, A.: "A New Method of Measuring Dielectric Loss," *Journal of Applied Physics*, 1946, **17**, p. 160.
(15) Barlow, H. M., and Cullen, A. L.: "Microwave Measurements" (Constable, 1950).

(9) APPENDIX

Theory of Method used to Measure Refractive Index and Normalized Impedance

Many artificial-dielectric media have relative permeabilities different from unity and so the method of measurement due to Roberts and von Hippel[14] cannot be used. It may, however, be extended,[15] the procedure being to determine the impedances Z_{SC} and Z_{OC} at the front surface of a slab of the medium of length l when the rear surface is terminated by a short-circuit and an open circuit respectively. If S_s and S_o are the measured standing-wave ratios under these conditions, and x_s and x_o are the corresponding distances between a field minimum and the front surface, then

$$Z_{SC} = Z_1 \frac{S_s - j\tan\beta_0 x_s}{1 - jS_s \tan\beta_0 x_s}$$

with a corresponding expression for Z_{OC}. Z_1 is the characteristic impedance of the empty line and β_0 (equal to $2\pi/\lambda_0$) is the free-space phase-change coefficient. The propagation constant γ and the characteristic impedance Z of the line, filled with the medium are then given by

$$\tanh \gamma l = (Z_{SC}/Z_{OC})^{\frac{1}{2}} : Z = (Z_{SC}Z_{OC})^{\frac{1}{2}}$$

The simplest case is that of a loss-free medium, when the two values of S become zero and γ reduces to the imaginary quantity $j\beta$, β being the phase-change coefficient of the medium. Then

$$\tan \beta l = (\tan\beta_0 x_s \cot\beta_0 x_o)^{\frac{1}{2}} : Z = Z_1(\tan\beta_0 x_s \tan\beta_0 x_o)^{\frac{1}{2}}$$

The ambiguity associated with the multiple roots of the first equation may always be resolved by making measurements on more than one length of sample.

Errors in β and Z arise because of errors in the measured values of x_o and x_s: these are most serious if either $\beta_o x_o$ or $\beta_o x_s$ is an odd multiple of $\pi/2$ since the tangent is then changing most rapidly. A small error under this condition may even change the sign of one of the tangents leading to imaginary values for the square roots.

621.372.2:621.315.61

Paper No. 1701 R
Jan. 1955

A GENERAL EXPERIMENTAL METHOD TO DETERMINE THE PROPERTIES OF ARTIFICIAL MEDIA AT CENTIMETRE WAVELENGTHS, APPLIED TO AN ARRAY OF PARALLEL METALLIC PLATES

By R. I. PRIMICH, Ph.D., B.Sc., Graduate.

(*The paper was first received* 28*th January, and in revised form* 11*th May*, 1954.)

SUMMARY

This paper describes measurements at centimetre wavelengths (8·0–11·0cm) on a system representing a semi-infinite array of infinitely thin, perfectly conducting, metallic plates. The purpose is to illustrate a general experimental method of determining the properties of such artificial media. The parallel-plate system was chosen since, in addition to exhibiting the interface phenomena peculiar to artificial media, it is the only one for which a rigorous solution is known. The measurements were made using a parallel-strip transmission-line[1] specially developed for the study of two-dimensional structures which are non-uniform transverse to the direction of propagation.

LIST OF SYMBOLS

Z_0 = Characteristic impedance of the H_{01} mode, relative to the free-space value.
ρ = Magnitude of electric field reflection-coefficient.
$\overline{r_{ab}}$ = Complex electric field reflection-coefficient, between media a and b with the incident wave in medium a.
$\overline{t_{ab}}$ = Complex transmission-coefficient for the wave transmitted into medium b with the incident wave in a.
Φ, ϕ = Phase of complex reflection-coefficients.
θ = Phase of complex transmission-coefficients.
R = Magnitude of reflection coefficient for a finite slab.
ψ = Phase of reflection coefficient for a finite slab.
l_p = Length of metallic plate in a slab of a parallel-plate medium.
q = Peak to peak value of the Weissfloch curve.
n = Refractive index of the parallel-plate medium.
λ_0 = Free-space wavelength.
λ_p = Plate-medium wavelength.

(1) INTRODUCTION

Fig. 1 illustrates an array of semi-infinite, infinitely thin, perfectly conducting, metallic plates. The plate spacing a, is so chosen that for a plane wave incident in the plane of the paper and of polarization as shown, only the H_{01} mode can propagate freely between the plates. A rigorous solution (usually referred to as the Carlson–Heins solution), which gives the total field at any point, is known.[2,3,4] This field consists of a freely-propagating reflected plane wave in free space, the transmitted wave in the plate region (H_{01} mode), and an infinite number of evanescent modes in the region of the physical interface. To obtain the reflected (or transmitted) wave, it is assumed that the field point is so far distant from the interface that the evanescent modes are of negligible amplitude.

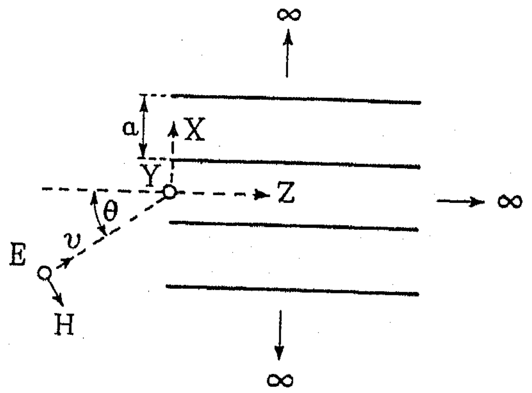


Fig. 1.—Semi-infinite array of parallel plates.

The information that is usually required (for microwave lens applications in particular), is the refractive index n of the plate system and the amplitudes of the transmitted and reflected waves. It is found that the electric field reflection and transmission coefficients, referred to the interface, are complex owing to energy storage in the evanescent fields.[2] There is, however, one simplification. If Z_0 is the normalized characteristic impedance of the H_{01} mode,[5] and ρ is the magnitude of the reflection coefficient, then $\rho = \frac{Z_0 - 1}{Z_0 + 1}$, as would be expected for a homogeneous material of impedance Z_0.

This analogy is incomplete since it ignores the phase shifts at the interface. Furthermore, it is a special case, for if the plates are of finite thickness, with the spacing between the plates the same as in an infinitely thin plate system, Z_0 is unchanged, but ρ is no longer given by $\rho = \frac{Z_0 - 1}{Z_0 + 1}$. In general, in any artificial medium in which Z_0 is clearly definable, it cannot be assumed that ρ is found from $\frac{Z_0 - 1}{Z_0 + 1}$. This is due to the presence of evanescent fields in the interface region which are necessary to match the free-space field to that in the medium. As the solution of the boundary-value problem for the semi-infinite structure has so far been obtained for only one special case,[2] the importance of the experimental determination of the behaviour of the interface is obvious.

The following Section will discuss the specification of the interface properties, which, together with the refractive index, completely describe an artificial medium.

(2) THE PROPERTIES OF AN INTERFACE BETWEEN FREE SPACE AND AN ARTIFICIAL MATERIAL

There are various ways in which the properties of an interface between free space and an artificial medium may be specified.

In one method, the interface region is replaced by a material whose refractive index (or permittivity and permeability) varies in some manner from the free-space value to that of the infinite artificial medium.[1,6,7] This idea has been used by Rice[8] and Piloty[9] in waveguide problems.

A different point of view is arrived at by considering one interface of an artificial material in a practical application. The reflected and transmitted fields at some distance from this

---

Written contributions on papers published without being read at meetings are invited for consideration with a view to publication.

Dr. Primich is in the Electrical Engineering Department of Imperial College of Science and Technology, University of London.

interface are then of interest, and these can be completely specified by the electric field reflection and transmission coefficients referred to any known plane. The physical interface is an obvious and convenient choice, which will lead, in general, to complex quantities. As the evanescent fields are of negligible amplitude, at a distance greater than one wavelength from the interface, the total field at any point beyond this region may be taken to be that of the main propagating modes, and to this extent the above representation is exact. It may be imagined that the effect of the evanescent fields, which in practice are limited in extent, has been replaced by a discontinuity specified by the complex coefficients. Immediately to either side of this supposed discontinuity are uniform semi-infinite regions of either free space or the material. A knowledge of the refractive index of the material then completes the amount of information which is necessary to describe the practical behaviour of a semi-infinite artificial medium.

A completely equivalent representation is that of the junction between two transmission lines in which reference planes for the reflected and transmitted waves are specified.[10] The transformation to complex reflection and transmission coefficients is straightforward.

The use of complex reflection and transmission coefficients is inherent in the Carlson–Heins formulation of the problem,[2] but Lengyel[11] has discussed this method in detail with particular reference to parallel plates, and has obtained relations of a general nature which will be outlined below.

$\overline{r_{ab}}$, $\overline{r_{ba}}$, $\overline{t_{ab}}$ and $\overline{t_{ba}}$ are the complex reflection and transmission coefficients at the interface (0,0), between two loss-free media a and b (Fig. 2). The order of suffixes a and b indicates the sense of the incident wave, e.g. $\overline{r_{ab}}$ is the wave reflected back into a for a unit wave incident on (0,0) from medium a to b.

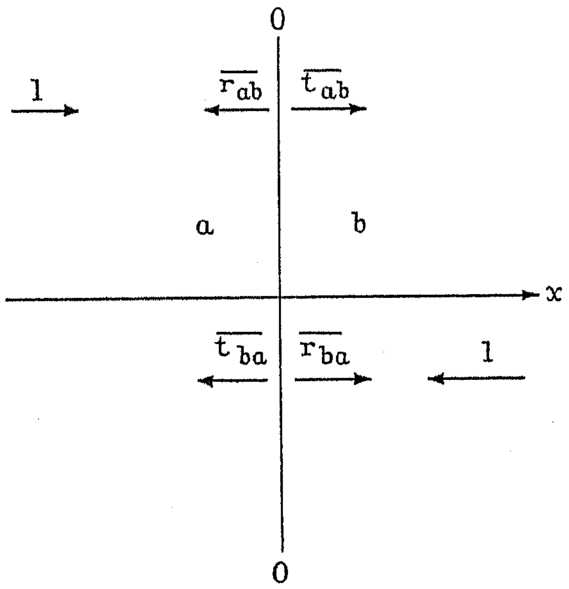


Fig. 2.—Reflection and transmission coefficients at the surface between two media.

$\overline{t_{ab}}$ is the wave transmitted into b for unit wave incident on (0,0) from medium a.

The phase of a reflection coefficient will be denoted by ϕ.

The phase of a transmission coefficient will be denoted by θ.

$$\text{i.e. } \overline{r_{ab}} = \rho_{ab} \exp(-j\phi_{ab}) \qquad \overline{r_{ba}} = \rho_{ba} \exp(-j\phi_{ba})$$

$$\overline{t_{ab}} = \tau_{ab} \exp(-j\theta_{ab}) \qquad \overline{t_{ba}} = \tau_{ba} \exp(-j\theta_{ba})$$

All phase angles are referred to (0,0) in Fig. 2, but by changing this reference plane the generality of the relations to follow may be extended.

$\overline{r_{ab}}$ and $\overline{t_{ba}}$ are now referred to some plane x_1, and $\overline{r_{ba}}$ and $\overline{t_{ab}}$ to some plane x_2 and they thus become the external coefficients of some medium of propagation length $x_1 + x_2$ (Fig. 3). The relations to be given will then apply to a single interface separating two media or to the two interfaces bounding a finite length of a medium.

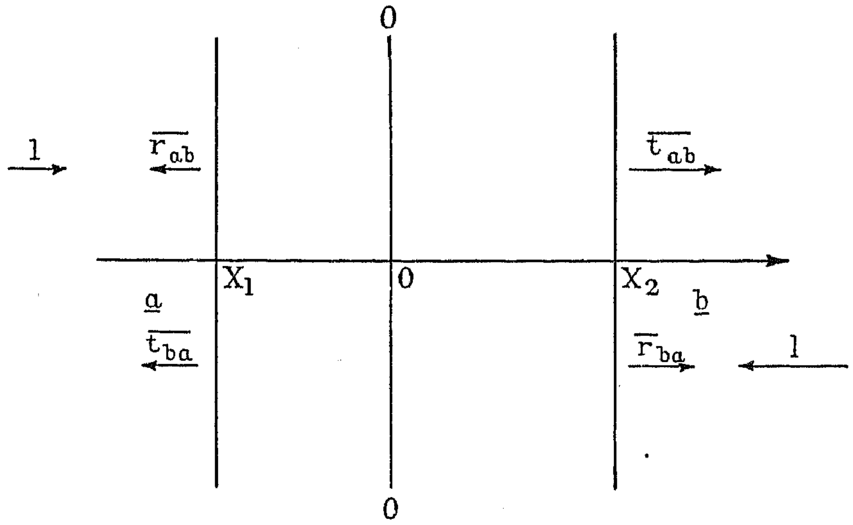


Fig. 3.—Reflection and transmission coefficients referred to two different planes.

Evanescent modes in the region of the interface are permissible, and the coefficients given above will then describe a single propagating mode on either side of the interface.

On the assumption of no loss and in conjunction with the reciprocity conditions, one has:

$$|\overline{r_{ab}}| = |\overline{r_{ba}}| = \rho \text{ say}$$

$$|\overline{t_{ab}}| = |\overline{t_{ba}}| = \tau \text{ say}$$

$$1 - \rho^2 = \tau^2$$

$$\text{and} \quad \left.\begin{aligned} \arg \overline{t_{ab}} &= \arg \overline{t_{ba}} = \theta \\ \phi_{ab} + \phi_{ba} - 2\theta &= \pm \pi \end{aligned}\right\}$$

$$\text{where} \quad \phi_{ab} = \arg \overline{r_{ab}} \text{ and } \phi_{ba} = \arg \overline{r_{ba}}$$

Consequently only three quantities, ρ or τ and two of the phase angles ϕ_{ab}, ϕ_{ba} or θ are required to specify the properties of an interface.

(3) EXPERIMENTAL

(3.1) Survey of the Literature

The majority of the work on parallel-plate media has been carried out with the object of confirming the Carlson–Heins theory. Techniques normally employed in free space were used.

Cochrane[12] employed a composite arrangement of a plano-convex lens and a parallel-sided prism constructed of metal plates. He found good agreement for the angle of deviation and consequently for the refractive index, but not for the amplitude of the transmitted wave.

Epstein[13] used a null method, based on a free-space interferometer, to obtain the phase of the transmitted wave, and Lengyel[14] with the aid of a microwave interferometer, obtained the magnitude of the reflection coefficient and the phase of the transmission coefficient for a metal-plate specimen of finite length.

Ruze and Young[15] (who first used two equipotential planes, but with sending and receiving horns) and Brady[16] studied the reflected wave with total absorption of the transmitted wave.

In general, good agreement with the Carlson–Heins theory was obtained, and most of the discrepancy, apart from the experimental error, was attributed to thick plate effects. However there are two main criticisms.

In the first place, none of the above work was sufficiently fundamental, i.e. no attempt was made to measure all three of the interface parameters, as well as the refractive index. In the case of Lengyel's work good agreement would have meant a confirmation of the theory, but it is impossible to determine the error in each of the single-interface parameters. In those cases where single-interface parameters were measured, the amount of information obtained was insufficient.

Secondly, it is felt that the general accuracy and consistency was not sufficient to discriminate between experimental error and, for instance, thick-plate effects. An exception is the work done by Brady,[16] in which for large angles of incidence where the reflected power level was high, the agreement is satisfactory. However, for small angles of incidence this is no longer true.

In the following Sections a description will be given of an experimental technique by which the single-interface parameters and the refractive index can be determined and in which, it is believed, greater accuracy has been achieved.

(3.2) The Parallel-Strip Transmission Line

This is fully described elsewhere[1] and for the present may be regarded as an extension of waveguide techniques, in which the ordinary transmission-line relations are applicable, rather than as a device in which a controlled free-space-field is available. It should be pointed out that the finite width of the line introduces an important source of error in that the incident field is only an approximation to a uniform-plane wave of infinite extent.[1] The extent to which this assumption affects the accuracy will be discussed later.

A system of parallel plates with the electric vector parallel to the plate edges is a two-dimensional problem with variation transverse to the propagation direction. This prevents it being studied in a waveguide, but it is suitable for the transmission line provided that the number of plates contained in the width of the line is sufficiently large.

In practice two types of line termination are possible, each leading to a different experimental technique.

(3.2.1) **The Matched-Load Termination.** (Fig. 4.)

A study of the interference pattern in front of the medium will yield the complex reflection-coefficient which is sufficient to specify completely the properties of a symmetrical slab of the medium. By obtaining this information for various lengths of medium the quantities for a single interface can be determined.

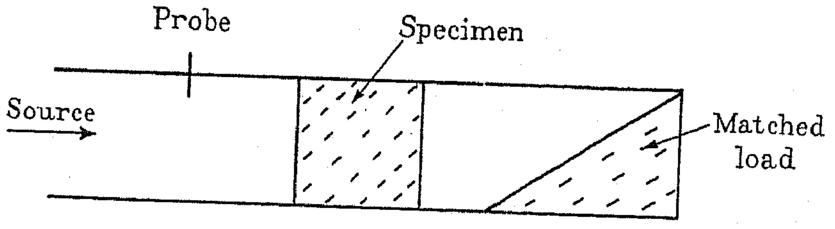


Fig. 4.—Matched-load method.

The main objections to this method are the need for continual calibration of the detecting system, and in particular for the strip transmission-line, the fluctuation of power level in the direction of propagation (due to Fresnel zone effects), which seriously limits the accuracy for small reflection coefficients.

The use of a single interface with the power transmitted into the plate region being totally absorbed, must be ruled out unless additional measurements are to be made on the transmitted wave.

(3.2.2) **Short-Circuit Termination.**

(3.2 2.1) *Single Interface.* (Fig. 5.)

By changing the length of the medium (lp) with respect to the short-circuit and obtaining the corresponding minimum position in the field pattern, the required parameters can be found.

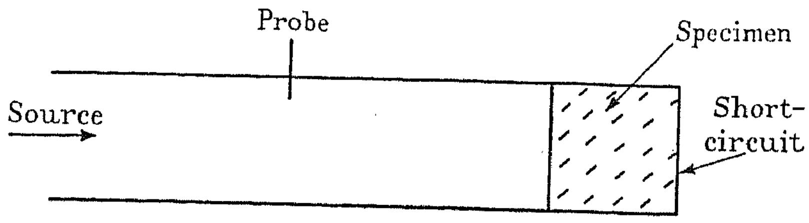


Fig. 5.—Short-circuit method with single interface.

This method proved to be unsatisfactory for the plate medium, since placing the specimen against the short-circuit on different occasions, gave a variation in minimum position of as much as two millimetres. The reason is thought to be imperfect contact between the plates of the medium and the short-circuit, which would give rise to a second set of evanescent modes (in the extreme case of no contact these would constitute a second interface), rather than cause a small error in length.

(3.2.2.2) *Two Interfaces.* (Fig. 6.)

The short-circuit is moved by known amounts relative to a slab of the medium of a given length. For each short-circuit position, the position of the minimum field strength (on the side of the specimen opposite to the short-circuit) is located. The curve relating the short-circuit position S, to the minimum position D (referred to any fixed point on the line), can be used to find the complex reflection-coefficient of the slab. Values of the latter for various slab lengths will give sufficient information for the single interface parameters to be found.

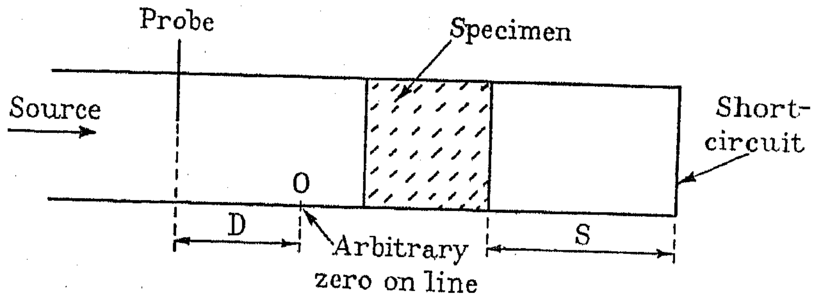


Fig. 6.—Short-circuit method with two interfaces.

The last two methods have the distinct advantage that they require only the determination of minima positions and absolute lengths. At 10cm this can be done with great accuracy, apart from any errors inherent in the transmission line. The reasons for these inherent errors are that the line is of finite width and that the measurements are being made in the Fresnel zone of the source, regarded as an aperture. The incident field is not uniform in the regions of the short-circuit,[1] and although the reflected field may be so (depending on the behaviour at the point of reflection), the resultant interference pattern will depart from the ideal and in effect will be the superposition of two waves of slightly different (and variable) wavelengths and amplitudes.

It was found by measurement that an actual minimum position and the distance between adjacent minima, could be in error by as much as 1%, with an average value, over a large number of measurements, of 0·4 to 0·5%. (Considering the line to be a large waveguide of width 50 cm would give a guide wavelength about 0·5% greater than the free-space value.) Non-uniformity, and curvature of the phase front of the incident wave, reflections from the sides of the line and diffraction by the specimen would all contribute to this error.

This error in minimum position appears almost directly in one of the phase angles of the reflection coefficient (since it is obtained from a distance measurement from the short-circuit), but appears indirectly in the magnitude of the same coefficient (since it is obtained from a minimum-position variation about a mean position). No satisfactory way of allowing for this error has been found and it was decided that its effect could best be determined by experiment.

The decision to use the short-circuit method with two interfaces was made on the basis of its simplicity, its accuracy compared to the matched-load technique (allowing for the error in minimum position which would be involved in either method), and the avoidance of the calibration of the detector crystal (apart from a check on the symmetry of the standing-wave pattern about the minimum position). A two-interface specimen, in preference to one interface, was chosen because of the contact difficulty mentioned previously. The details of the actual method are presented in Section 3.4.

(3.3) The Reference System

As a check on the accuracy of the transmission line and an illustration of the method of measurement, a detailed investigation is required of a medium, whose properties can be calculated. An array of thin metal plates, for which the Carlson–Heins theory is available, is suitable, and the plates can be made from tin foil supported by a low-loss dielectric. The Carlson–Heins solution applies only when the plates are embedded in a uniform medium of infinite extent, and to simulate this the arrangement of Fig. 7 is used. The dielectric is embedded beyond the plate interface, (0′,0′), by a distance l_s, sufficiently large to ensure that the evanescent field has negligible amplitude at (0,0). Section 11.2 shows how the reflection and transmission coefficients for a single effective interface between free space and the region of the plates can be calculated from the coefficients for the interfaces (0,0) and (0′,0′). The coefficients for (0,0) are calculated for the measured value of the relative permittivity of the dielectric and those for (0′,0′) from the Carlson–Heins theory. A comparison of the calculated and measured values of the coefficients for the single effective interface therefore suffices as a check on the proposed measurement technique.

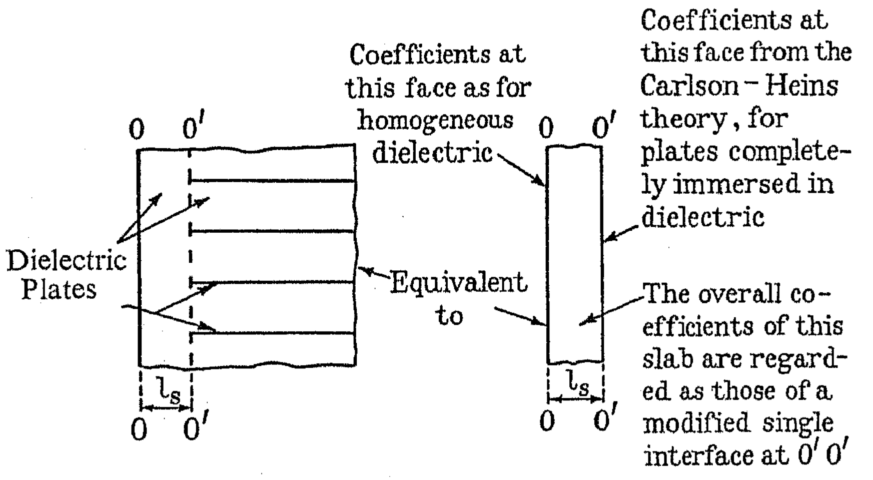

0
0′
Dielectric
Plates
l_s
0
0′
Coefficients at this face as for homogeneous dielectric
Equivalent to
0
0′
l_s
0
0′
Coefficients at this face from the Carlson-Heins theory, for plates completely immersed in dielectric
The overall coefficients of this slab are regarded as those of a modified single interface at 0′ 0′

Fig. 7.—Application of the Carlson–Heins solution to a system with a dielectric projection.

(3.4) The Experimental Method (Fig. 8)

For reasons given previously it was decided to use a slab of the plate medium with a short-circuit termination to the line.

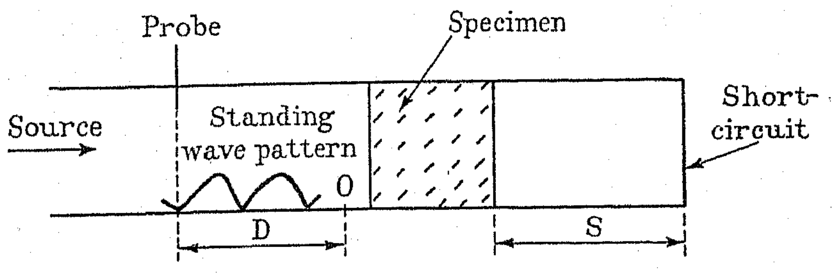


Fig. 8.—Short-circuit method with two interfaces.

The measurements consist quite simply of locating a minimum position on one side of the medium for various positions of the short-circuit relative to the other interface of the medium. This type of curve relating a short-circuit position to a minimum

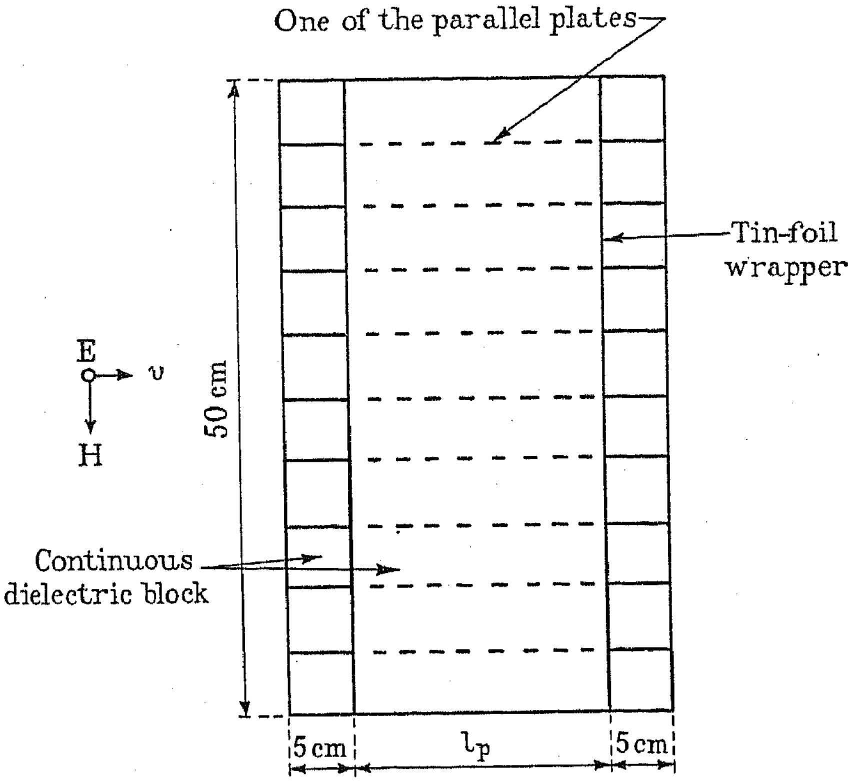


Fig. 9.—Parallel-plate specimen with dielectric projection.

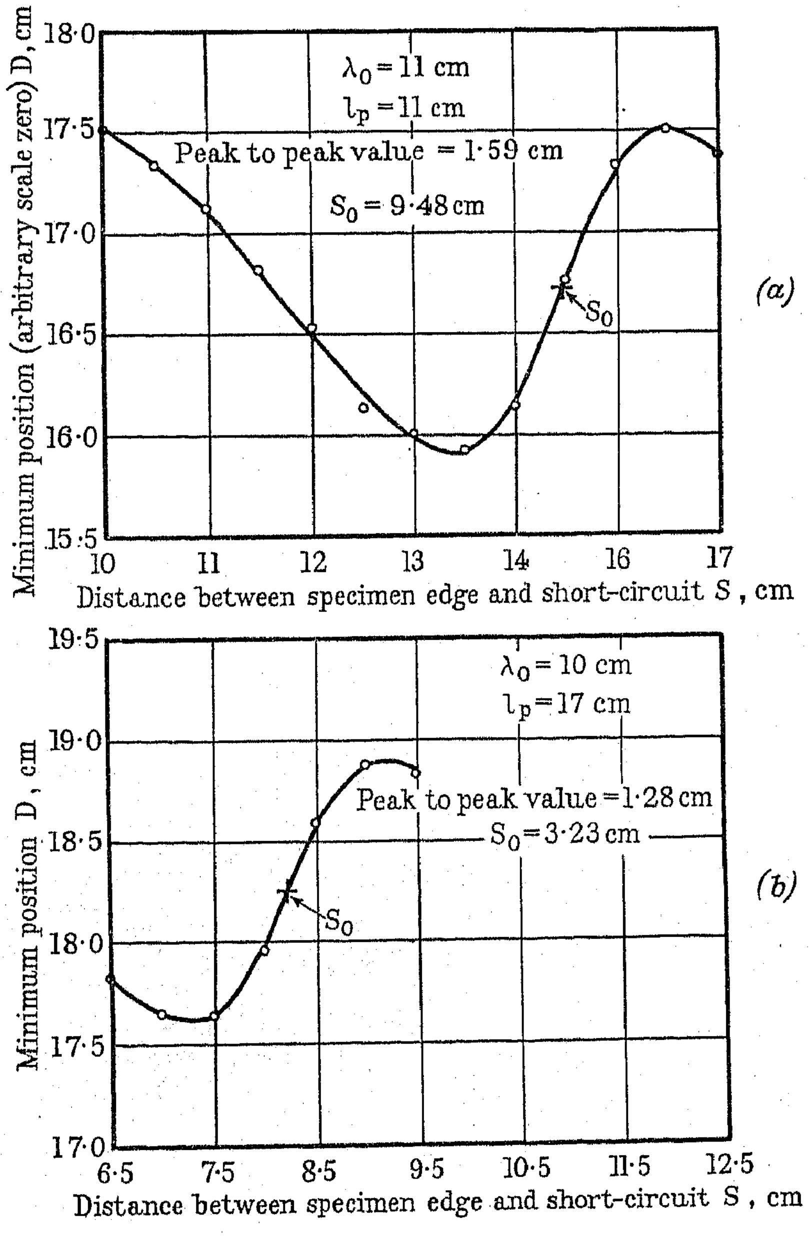


Fig. 10.—Typical Weissfloch curves.

position on either side of an obstacle was first discussed by Weissfloch[17] who pointed out that it could be used to determine the properties of a four-terminal network (or a discontinuity) in a transmission line and discussed it in terms of the transforming properties of such a network.

Others have used similar curves and have interpreted them in different ways. Oliver[18] and Marcuvitz[10] have derived equivalent circuits and Feenberg[19] and Barlow and Cullen[21] have been interested in reflection coefficient and standing-wave ratio.

In the method, the specimen, assembled as in Fig. 9, was made to fit tightly between the plates of the line and was located with respect to the short-circuit by wedging the plates apart, fitting machined spacers to obtain the distance S, removing the wedges and then the spacers. The corresponding minimum position was then located. The absolute value of D is not required, and all that is necessary is the relative movement of the minimum position (on some scale attached to the line) as S is varied. S was varied by moving the specimen through half-centimetre intervals. Fig. 10 shows typical curves for half-wavelength and quarter-wavelength changes in S. The latter change is sufficient since the peak value gives the magnitude of the reflection coefficient for the particular slab of the medium being measured. The phase follows from $S = S_0$, the point midway between the peak values (see Section 4).

Repeating this curve at various wavelengths, for a fixed medium length, yielded two curves relating the magnitude R and phase ψ of the reflection coefficient of the slab, to the wavelength (Figs. 11 and 12). Similar sets of curves were obtained for half-centimetre changes in the specimen length.

These curves were then replotted for fixed wavelengths to give the magnitude R and phase ψ of the reflection coefficient, in terms of the medium length l_p (Figs. 13 and 14). The parameters for the modified single interface were obtained from the latter curves (see Section 4).

(3.5) Description of the Specimen

The dielectric material was expanded ebonite of permittivity 1·07, and this was machined in blocks of width equal to the plate spacing in the parallel-plate medium. The height of each

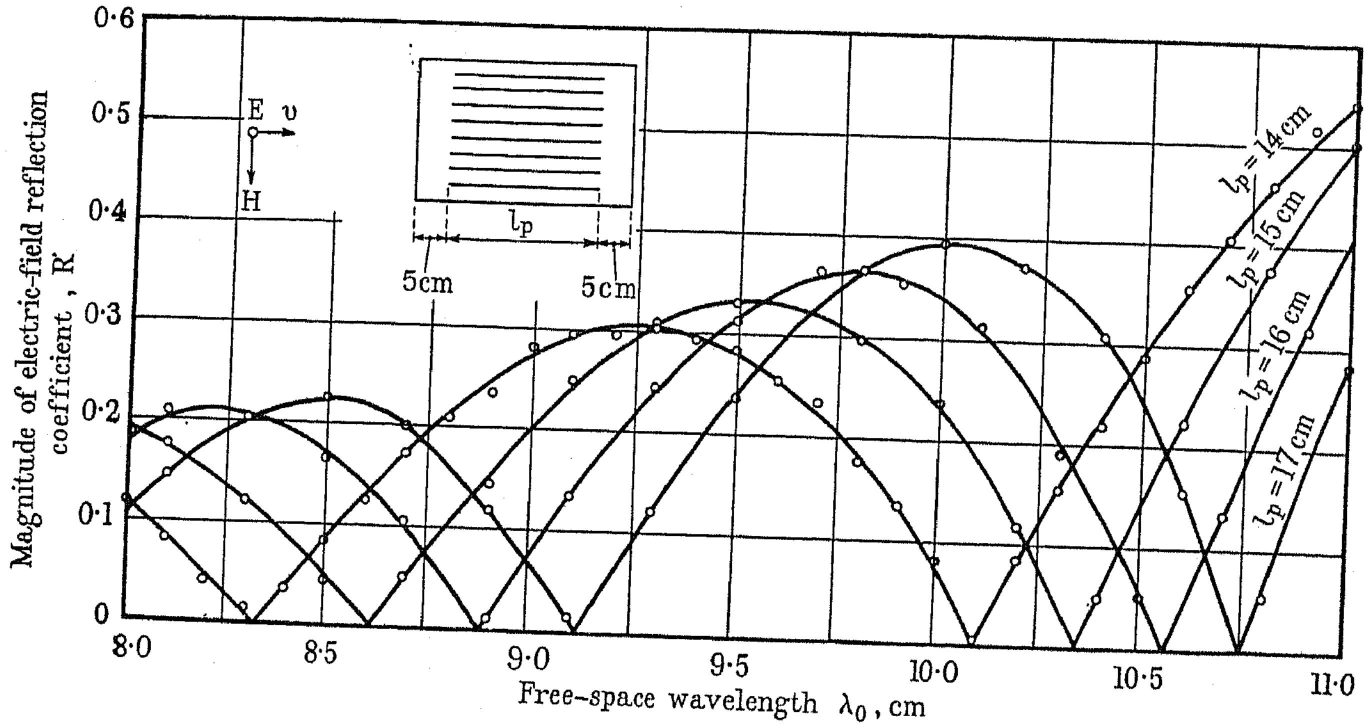


Fig. 11.—Magnitude of reflection coefficient against wavelength.

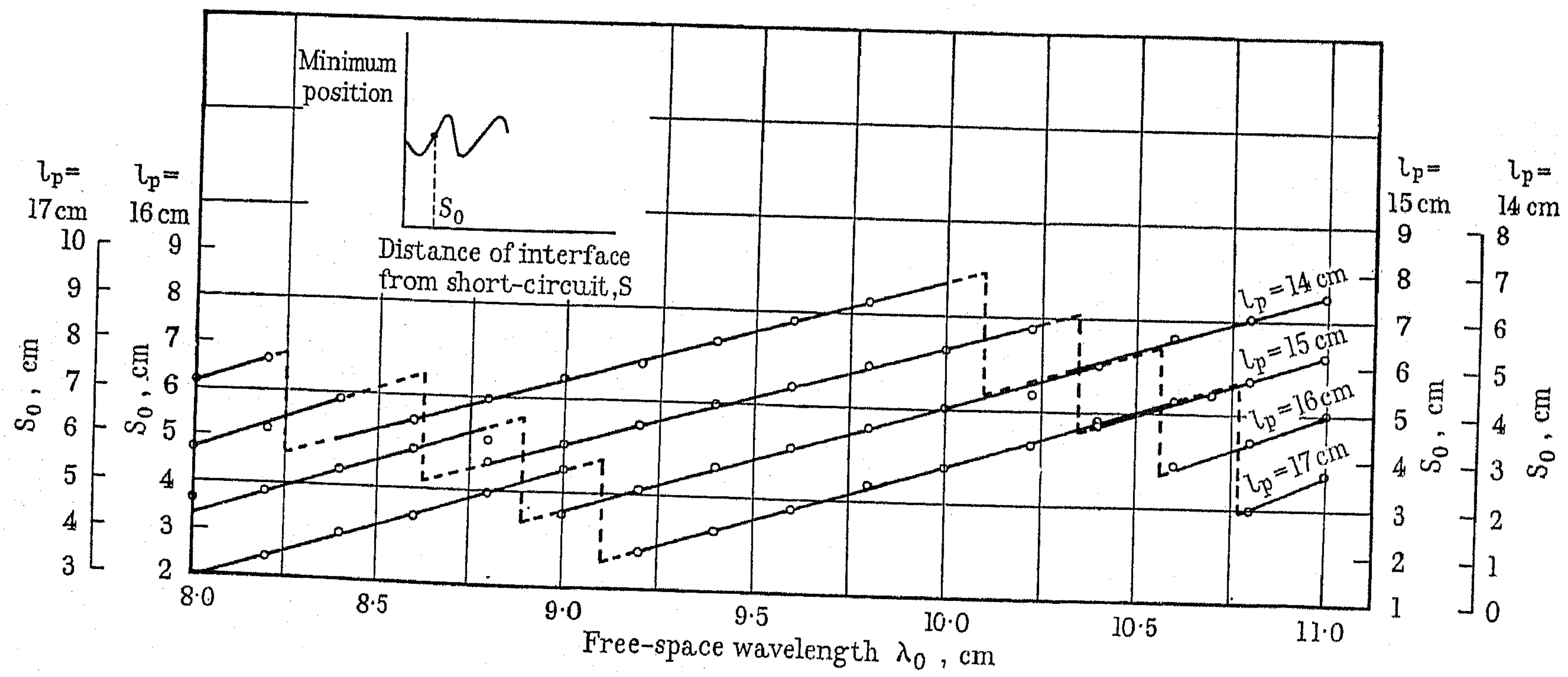


Fig. 12.—Phase of reflection coefficient against wavelength.

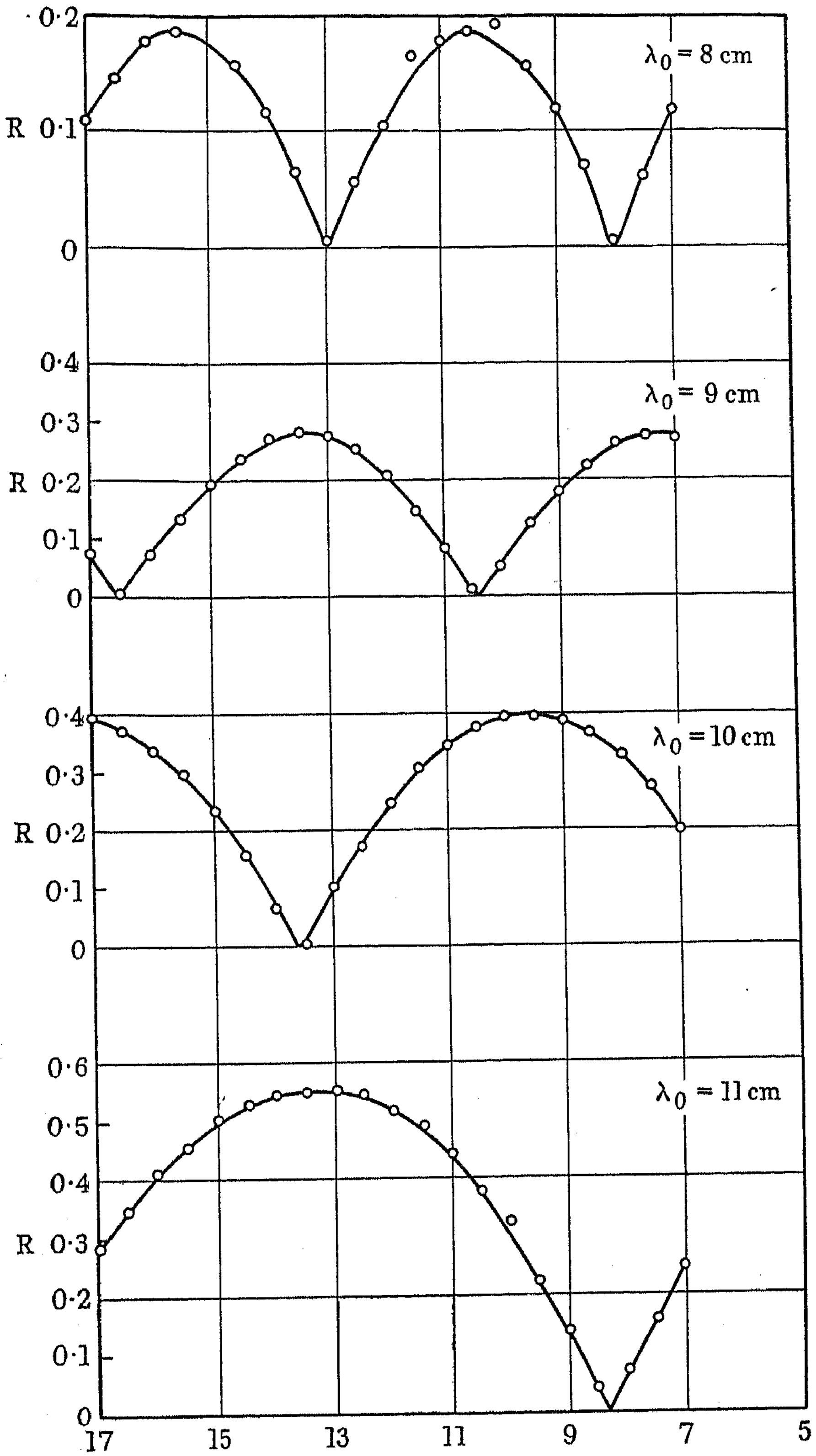


Fig. 13.—Magnitude of reflection coefficient against plate length (l_p).

block was equal to the spacing between the conductors in the transmission line (3 cm), and the length of each block was the same as the length of the sample. The parallel plates of the specimen were formed by wrapping 0·001 in tin foil around each block with an overlap on the broad side adjacent to a line conductor. Eight blocks were placed side by side to give nine parallel plates each 0·002 in thick. The dielectric interface-slabs were obtained by making the tin foil 5 cm shorter than the block length, at each end. The whole specimen was held together by wrapping it in a large sheet of tin foil (covering the length of each plate only), and gluing a small part of the overlap with polystyrene. Judging by the consistency and reproducibility of the results (the specimen was dismantled and re-assembled several times), the contact problem was satisfactorily solved by tight wrapping and a tight fit in the line, and by having all discontinuities in the foil transverse to the direction of propagation. To reduce the specimen length it was necessary to dismantle it, but this form of construction ensured that the mechanical accuracy and electrical performance were within the limits desired.

The permittivity of the ebonite was measured in the transmission line from 8·0–11·0 cm, and the results were consistent to within ±1/3%. The absolute value agreed with those from waveguide measurements and at low frequencies. The accuracy is attributable to the long specimens used (30 cm).

For an ideal specimen of finite length, accurately constructed and possessing no loss, the reflection coefficient should be zero at some frequency. It was found that the smallest measurable reflection coefficient was of the order of 0·003 (±0·001) which clearly indicates that any error arising from the above causes is insignificant.

The dimensions of the specimen are given in Fig. 9.

(4) THE INTERPRETATION OF RESULTS

(4.1) The Weissfloch Curve

If q is the peak-to-peak value of the curve (Fig. 10), then[21]

$$R = \sin (2\pi q/2\lambda_0)$$

= magnitude of reflection coefficient for the slab

and $$\psi = 2\pi 2S_0/\lambda_0$$

= phase of reflection coefficient for the slab.

These are plotted in terms of wavelength (Figs. 11 and 12), or specimen length (Figs. 13 and 14). The theoretical equations for R and ψ are:

$$R = \frac{2P \sin \xi}{\sqrt{[(1 - P^2)^2 + (2P \sin \xi)^2]}} \quad \text{(See Sections 11.1 and 11.2)}$$

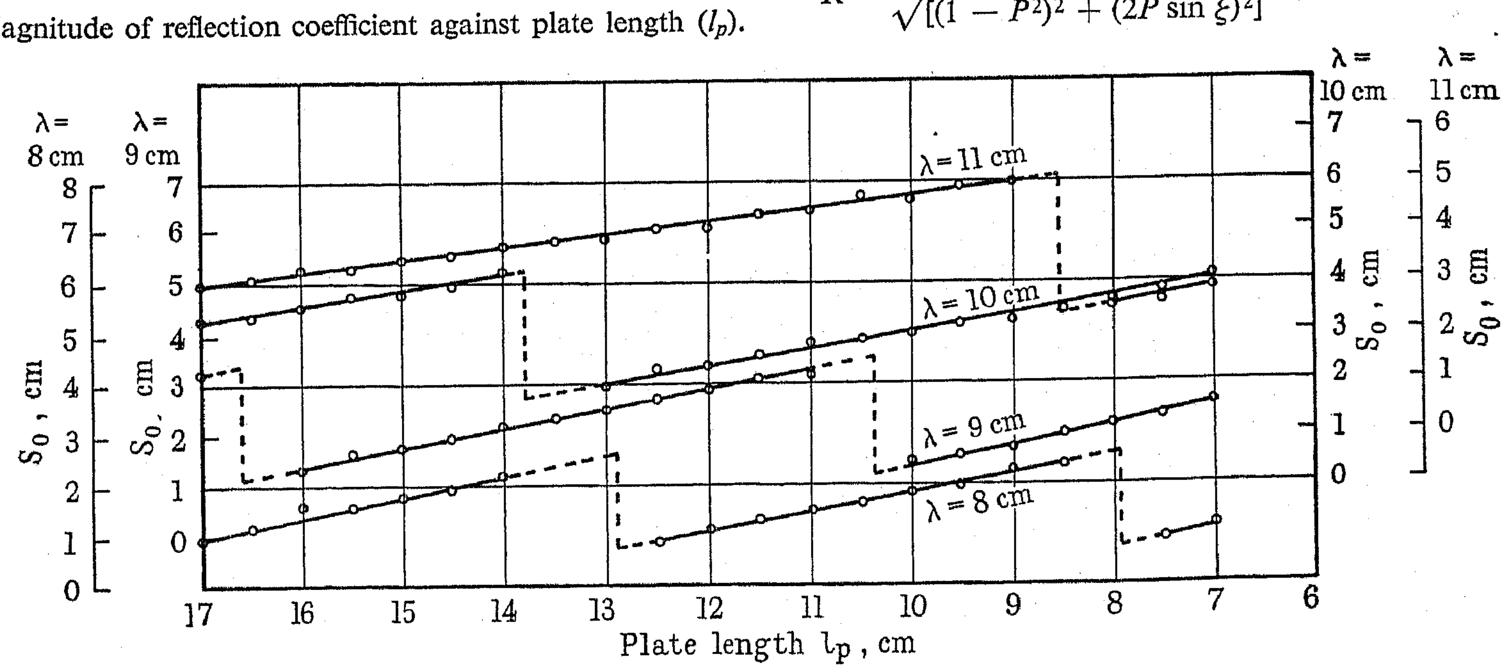


Fig. 14.—Phase of reflection coefficient against plate length (l_p).

$$\psi = \Phi_{11} - \arctan\left[\frac{(1-P^2)}{(1+P^2)\tan\xi}\right]$$

$$\xi = (2\pi n l_p/\lambda_0) + \Phi_{22}$$

P = magnitude of reflection coefficient of the modified single interface.

Φ_{11} = phase of reflection coefficient (free-space side), at the plate edges.

Φ_{22} = phase of reflection coefficient (plate-medium side) at the plate edges.

(4.2) Determination of P, Φ_{11}, Φ_{22}

R (as a function of l_p) will have maxima at $\xi = m\pi/2$, $m = 1, 3, 5, \ldots$, and $R_{max} = 2P/(1 + P^2)$, which determines P.

The period of $R = f(l_p)$ is $\lambda_p/2$ and so $n = \lambda_0/\lambda_p$.

At $\xi = 0$, $R = 0$ and at this point let $l_p = l_{p0}$.

Therefore $\quad \xi = \Phi_{22} + (2\pi n l_{p0}/\lambda_0) = 0$

or $\quad \Phi_{22} = -(2\pi n l_{p0}/\lambda_0)$

From $\quad \psi = \Phi_{11} - \arctan\left[\dfrac{(1-P^2)}{(1+P^2)\tan\xi}\right]$

$\psi = \Phi_{11}$, when $\arctan\left[\dfrac{(1-P^2)}{(1+P^2)\tan\xi}\right] = 0$

or $\quad \xi = \dfrac{\pi}{2} \ldots$

i.e. $\quad l_p = l_{p0} + \lambda_p/4$

From the curve $S_0 = f(l_p)$, at $l_p = l_{p0} + \lambda_p/4$, $S_0 = \left(\dfrac{1}{2}\dfrac{\lambda_0}{2\pi}\right)\psi$

Therefore $\quad \Phi_{11} = \psi = 2S_0 2\pi/\lambda_0$ at $l_p = l_{p0} + \lambda_p/4$

(4.3) Determination of n

$$n = \lambda_0/\lambda_p$$

where $\lambda_p/2$ is the periodicity of the curve $R = f(l_p)$.

This result was used in Section 3.2 for the determination of Φ_{22}.

In the experimental determination of Φ_{11} and Φ_{22}, there is an uncertainty of π in the absolute value which can be resolved as follows:

Barlow and Cullen[21] show that the absolute value of Φ_{11} is found from the mean point on the Weissfloch curve, in the region of minimum slope, nearest to the short-circuit ($S = 0$). The curves that have been plotted give S_0 in the region of maximum slope, so that it is necessary to add (or subtract) a quarter-wavelength and then to ascertain that this point is the one nearest to the short-circuit (in comparison with other points of minimum slope).

If l_{p0} is chosen to be the shortest length at which $R = 0$, i.e. from any l_{p0} subtract $\dfrac{m\lambda_p}{2}$, $m = 1, 2, \ldots$, then Φ_{22} will have its correct absolute value.

These formulae are best illustrated by a specific example, say at $\lambda_0 = 10$ cm.

From Fig. 10, $l_p = 17$ cm, $q = 1{\cdot}28$ cm.

$$\begin{aligned} \text{Therefore}\quad R &= \sin(2\pi \times 1{\cdot}28/2 \times 10) \\ &= 0{\cdot}392 \\ S_0 &= 8{\cdot}23\,\text{cm (maximum-slope region)} \\ \text{or}\quad & 8{\cdot}23 - 10/4 = 5{\cdot}73\ \text{(minimum-slope region)} \end{aligned}$$

These values are then plotted on the curves $R = f_1(\lambda_0)$, $S_0 = f_2(\lambda_0)$, for $l_p = 17$ cm and $\lambda_0 = 10$ cm (Figs. 11 and 12).

From Fig. 13, at $\lambda_0 = 10$ cm, periodicity $= \lambda_p/2 = 7{\cdot}65$ cm (average for whole curve).

$$n = (10/2)/7{\cdot}65 = 0{\cdot}654$$

From Fig. 13, $\lambda_0 = 10$ cm.

$$R_{max} = 0{\cdot}398 = (2P)/(1 + P^2)$$

therefore $\quad P = 0{\cdot}208$

Also $R = 0$ at $l_{p0} = 13{\cdot}59$ cm (Fig. 13).

The first zero occurs at $l_{p0} = 13{\cdot}59 - 7{\cdot}65 = 5{\cdot}94$ cm.

$$\begin{aligned} \Phi_{22} &= (-2\pi n l_{p0}/\lambda_0) \\ &= (-2\pi \times 0{\cdot}654 \times 5{\cdot}94/10)\ \text{radians} \\ &= -139{\cdot}96\ \text{degrees} \end{aligned}$$

And $\quad \xi = \pi/2$ at $l_p = l_{p0} + \dfrac{\lambda_p}{4} = 5{\cdot}94 + \left(\dfrac{1}{4}\dfrac{10}{0{\cdot}654}\right) = 9{\cdot}76$ cm.

From Fig. 14, at $l_p = 9{\cdot}76$, $S_0 = 3{\cdot}11$ cm or $3{\cdot}11 - \dfrac{10}{4} = 0{\cdot}61$ (minimum-slope region).

$$\Phi_{11} = \psi = (-2\pi 2S_0/\lambda_0) = (-2\pi \times 1{\cdot}22/10) = -44$$

This value is referred to the front face of the dielectric, whereas it is required at the plate edges. The simplest way of changing

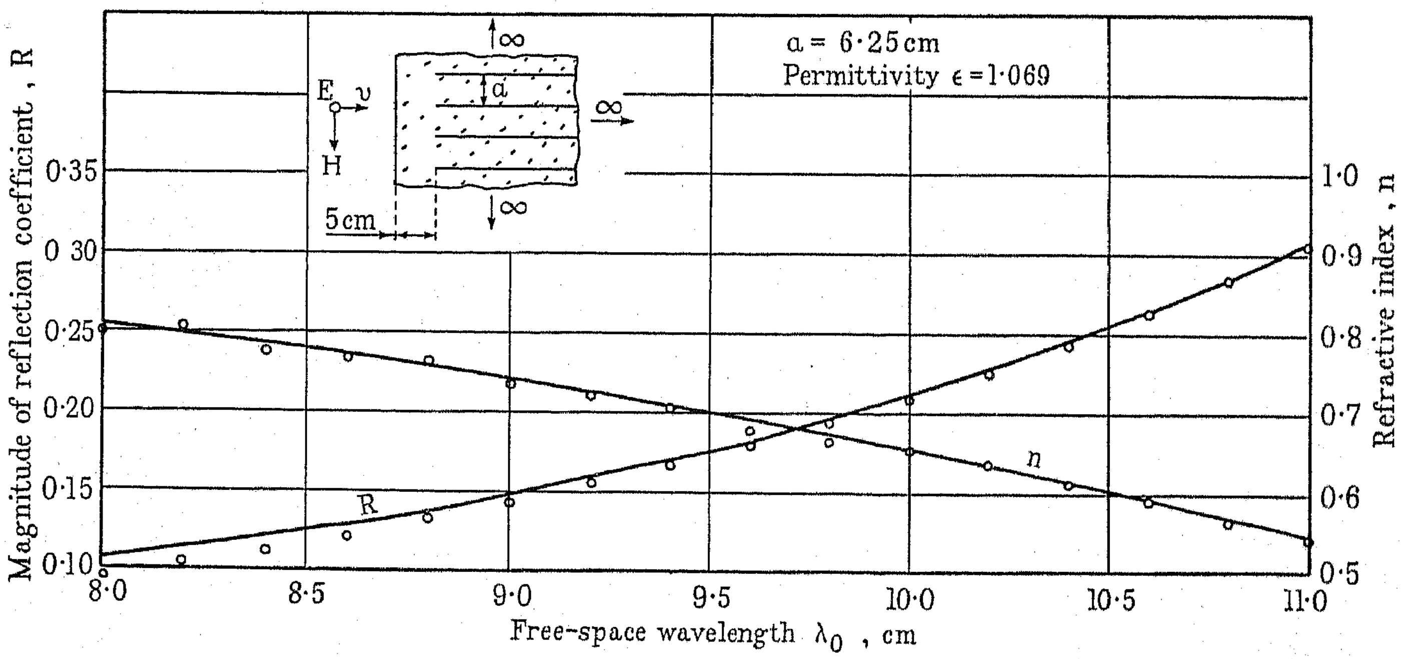


Fig. 15.—Magnitude of reflection coefficient and refractive index against wavelength, for a single interface.

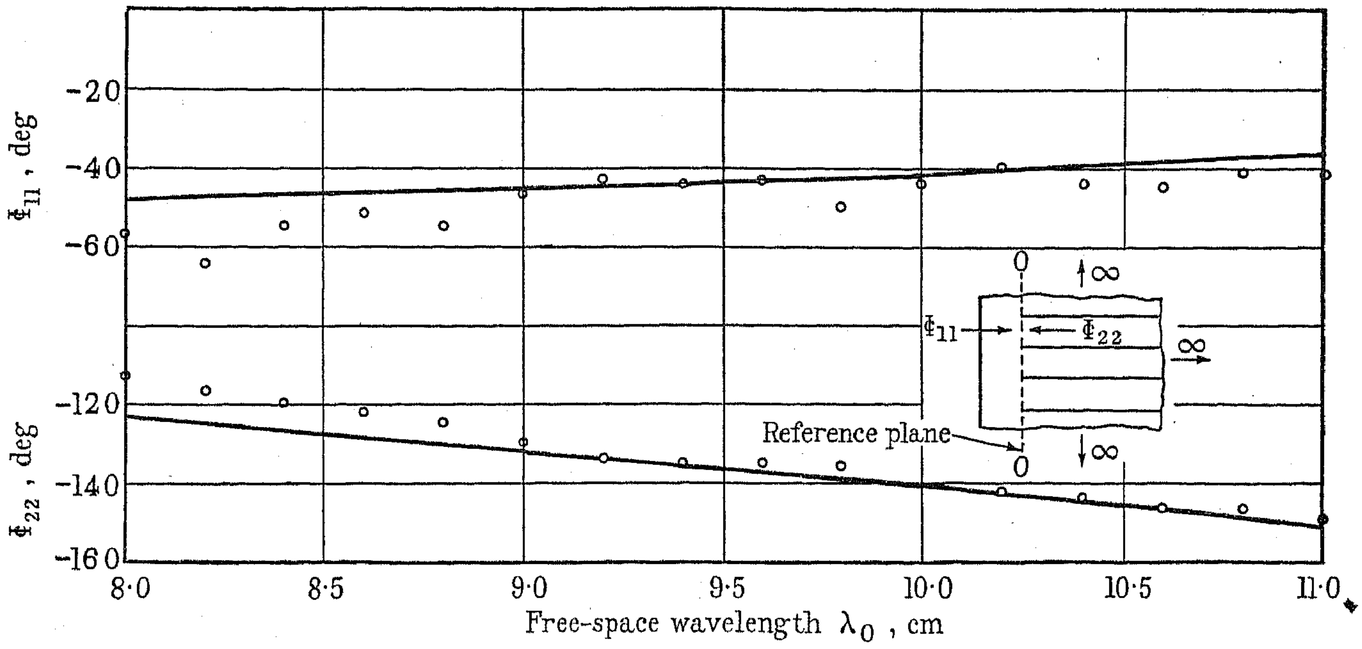


Fig. 16.—Phase of reflection coefficient for a single interface.

the reference plane is to add 5cm (length of the dielectric projection) to S_0.

i.e.
$$\Phi_{11} = [-2\pi 2(S_0 + 5)/\lambda_0]$$
$$= (-2\pi 2S_0/\lambda_0) - (2\pi 10/\lambda_0)$$

By coincidence this makes no difference at 10cm, but at other wavelengths it will have an effect.

(Note that the shift in reference plane takes place in free space and not in the dielectric.)

Figs. 14 and 15 give the final results for P and n and Φ_{11} and Φ_{22}. The derivation of the theoretical curves is given in Section 11.3.

(5) MEASUREMENTS ON SHORT PLATES

The purpose of these measurements was to determine for what length of plate, coupling between higher-order modes from the two interfaces became noticeable, i.e. the point at which the Carlson–Heins theory broke down. This is best studied by determining R as a function of l_p.

For no coupling between evanescent modes,

$$R = \frac{2P \sin \xi}{\sqrt{[(1 - P^2)^2 + (2P \sin \xi)^2]}}$$

whence, for $l_p = 0$, $\xi = \Phi_{22}$

$$R = \frac{2P \sin \Phi_{22}}{\sqrt{[(1 - P^2)^2 + (2P \sin \Phi_{22})^2]}}$$

which is still finite.

However, from physical reasoning, R must be zero when $l_p = 0$, so that the effect of coupling can be seen in the deviation of $R = f(l_p)$ from the theoretical curve

$$R = \frac{(2P \sin \xi)}{\sqrt{[(1 - P^2)^2 + (2P \sin \xi)^2]}}$$

Measurements on the previous specimen were continued for $\lambda_0 = 10{\cdot}0$ to $11{\cdot}0$cm and for l_p from 7 to 1cm. Fig. 17 shows the results for the worst case, i.e. $\lambda_0 = 11{\cdot}0$cm and $\lambda_p = 20{\cdot}4$cm.

The two curves agree up to $l_p = 5$cm, i.e. $l_p/\lambda_p = 1/4$, or about $\frac{1}{2}\lambda_0$. At $l_p = 3{\cdot}0$cm R has dropped from 0·56 to 0·525 or about 6%.

The conclusion is that the Carlson–Heins theory is applicable to slabs longer than a half-wavelength in free space. For quarter-wavelength slabs the error does not exceed 10%.

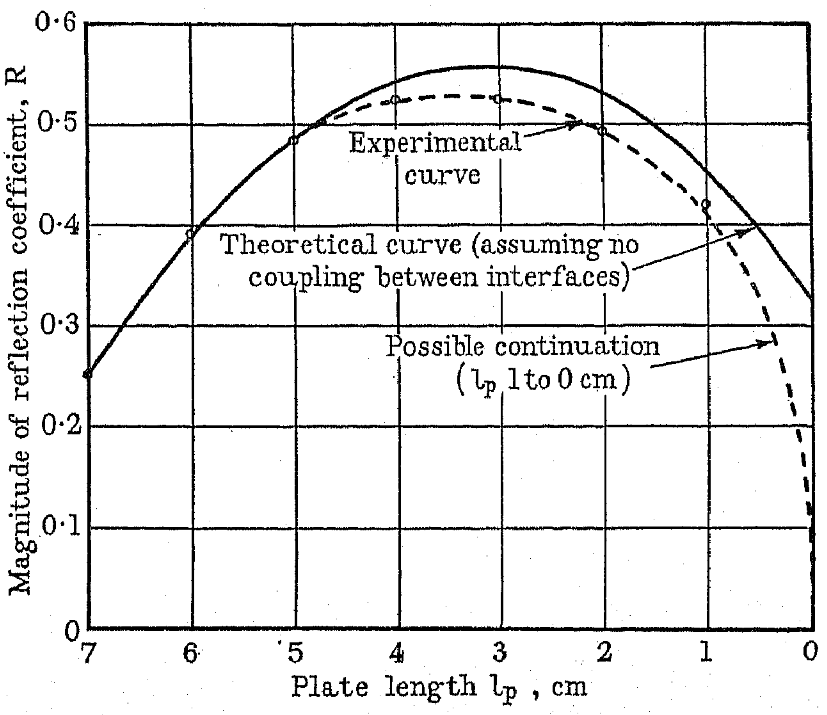


Fig. 17.—Magnitude of reflection coefficient for short plates.

Although these remarks apply to the magnitude of the reflection coefficient of a slab, it is expected that the phase would behave in a similar manner as the phase angles at a single interface contribute to R.

(6) DISCUSSION OF THE RESULTS

Fig. 15 shows that the results for the refractive index are satisfactory. The worst error is −2%, so that the accuracy can certainly be given to within ±2%. There is no definite evidence of a systematic error. The magnitude of the reflection coefficient does show a systematic error which increases with decreasing wavelength. The error at 9·0cm is −3% and becomes too severe at shorter wavelengths. The scatter appears to be about ±1%, so the resultant error could safely be specified as −1% to −5% of the theoretical value in the wavelength range 9·0–11·0cm.

Fig. 16 shows that the phase is within 3° from $\lambda_0 = 9{\cdot}0$cm to $\lambda_0 = 10{\cdot}0$cm, but there is greater deviation for wavelengths outside this range.

Consider the effect of mechanical spacing in contributing to the above errors. The actual spacing used in the calculations was arrived at by measuring the average specimen width between two straight edges and dividing by the number of spaces in the specimen. Because of the pressure exerted by the line on the specimen it might be thought that the actual spacing would increase beyond the measured value. To cause a 3% reduction in magnitude of reflection coefficient would require the overall width of the specimen to increase by 6mm, which is impossible. Similarly an error in the value of permittivity of the spacing dielectric would be too large to be admissible.

The presence of the interface slabs is a possible source of error in that the theory allowing for them may not be correct. As the theory is quite straightforward this is unlikely, and furthermore, measurements on similar systems in which these slabs were absent, showed that a systematic error of the same order persisted.

Although appreciable scatter was encountered for individual points (some 5% at the most), this is progressively reduced by the method employed, with the result that the final error is almost entirely of a systematic nature. The most likely origin of this is in the line itself. The fundamental cause is believed to be the finite width of the source. In conjunction with the finite width of the line conductors, it is almost impossible to calculate the actual behaviour of such a system and at the best it can only be discussed qualitatively.

One of the main consequences of using a source of finite size is that the region of practically uniform radiation is attained only beyond a certain distance from the source (i.e. beyond $2D^2/\lambda_0$, D being aperture width). For distances less than this (in the Fresnel region), the field departs from ideal, depending on the aperture size and the distance. The ideal would consist of an incident wave of some known amplitude distribution and a flat phase-front. A non-uniform amplitude distribution need not be regarded as a defect. In the transmission line, measurements were actually being carried out in the Fresnel region of the source and this is evident from results obtained under short-circuit conditions. The distance between adjacent minima gave an apparent wavelength which was variable and in error by 0 to +1%, with an average of $+\frac{1}{2}$%. The variable character arises from the nature of the incident field in the Fresnel region, which is not uniform, whereas the average error is connected with the taper of the field pattern. It might be thought that the latter error could be compensated for by treating the line as a large guide, or more simply by assuming that the measurements were made in a dielectric of permittivity 1/1·005. On the basis of such a correction the error is changed as follows:

	8·0 cm	9·0 cm	10·0 cm	11·0 cm
R (original) error ..	0	0	0	0
R (measured) ..	−8%	−3%	−3%	−1%
R (new)	−3%	$-1\frac{1}{2}$%	$+1\frac{1}{2}$%	+2%

If such a correction were used it would have to be a function of wavelength.

The situation is still further complicated by other factors. At the longer wavelengths the phase front of the incident wave showed considerable curvature,[1] which would certainly be detrimental. At the shorter wavelengths, radiation from the side of the line was noticeable, indicating the presence of unwanted reflections in the body of the line.

The finite width of the specimen itself could be troublesome in two ways. Diffraction from the sides of the specimen is not considered important as the field strength at the side of the line was found to be small. Of greater importance was the number of plates occurring across the width of the line. With the number used, measurements were being carried out well into the Fraunhofer zone of each plate spacing regarded as an aperture, but if the number is much further reduced the diffraction of each plate edge, rather than a plate medium, would be under consideration.

The foregoing remarks have merely served to show that there is a primary cause for a systematic error, the magnitude and direction of which approximate to that actually obtained, and that the situation is complicated by other factors none of which can be satisfactorily allowed for. As a result it has been found more convenient to use the above set of measurements to predict the error in later work.

(7) A COMPARISON OF THE SINGLE- AND DOUBLE INTERFACE METHODS

The main disadvantage of the method employed is that it consumes a great deal of time and labour and this must be weighed against the requirements of the work to be done. It appears to offer greatest accuracy, as measurements carried out on a single interface showed considerable scatter. The single-interface method is certainly quicker and easier to carry out and would be extremely useful for exploratory purposes, followed by more detailed measurements using two interfaces and yielding more specific results. For the transmission line in particular, the two-interface method has proved extremely useful since it is essentially an averaging process and would reduce scatter, which appears to have some support from the results. It should be pointed out that this method as applied to a parallel-plate medium is perfectly general and can be used for the study of any two-dimensional artificial structure.

(8) CONCLUSIONS

A method of specifying and determining the properties of an artificial medium has been demonstrated. The transmission line has made possible the accurate measurement of two-dimensional structures at centimetre wavelengths and its performance between 9·0 and 10·0cm is satisfactory within the following limits:

Magnitude of reflection coefficient (approximately 0·1): within 5% of the theoretical value (probably of the order of 3%);
Phase of reflection coefficient: within 5° and probably within 3°;
Refractive index: within 3%.

(9) ACKNOWLEDGMENTS

The author wishes to express his gratitude to Dr. Willis Jackson and Mr. John Brown for the guidance he has received from them and for their assistance in the preparation of the paper. He is indebted to Standard Telephones and Cables Ltd. for the supply of wide-band coaxial-line valves, which facilitated the measurements described.

(10) REFERENCES

(1) El-Kharadly, M. M. Z.: "Some Experiments on Artificial Dielectrics at Centimetre Wavelengths" (see page 17).
(2) Carlson, J. F., and Heins, A. E.: "The Reflection of an Electromagnetic Plane Wave by an Infinite Set of Plates," *Quarterly of Applied Mathematics*, 1947, **4**, p. 313, 1947, **5**, p. 82, and 1950, **8**, p. 281.
(3) Berz, Mrs. F.: "Reflection and Refraction of Microwaves at a Set of Parallel Metallic Plates," *Proceedings I.E.E.*, Paper No. 1038 R, January 1951 (**98**, Part III, p. 47).
(4) Whitehead, E. A. N.: "The Theory of Parallel-Plate Media for Microwave Lenses," *Proceedings I.E.E.*, Paper No. 1093 R, March 1951, (**98**, Part III, p. 133).
(5) Slater, J. C.: "Microwave Electronics" (van Nostrand, 1950), p. 8.
(6) Corkum, R. W.: "Isotropic Artificial Dielectrics," *Proceedings of the Institute of Radio Engineers*, 1952, **40**, p. 574.

(7) Susskind, C.: "Investigation of Obstacle-Type Artificial Dielectric," Ph.D. Thesis, Yale University, 1951.

(8) Rice, S. O.: "Reflections from Corners in Rectangular Waveguides—Conformal Transformation," *Bell System Technical Journal*, 1949, **28**, p. 104.

(9) Piloty, R.: "Das Feld in inhomogen Rechteckrohren bei anregung mit der H_{01}-Welle," *Zeitschrift für angewandte Physik*, 1949, I, **11**, p. 490.

(10) Marcuvitz, N.: "Waveguide Handbook," Radiation Laboratory Series, **10** (McGraw-Hill Book Co., 1951).

(11) Lengyel, B. A.: "A Note on Reflection and Transmission," *Journal of Applied Physics*, 1951, **22**, p. 263.

(12) Cochrane, C. A.: "An Experimental Verification of the Theory of Parallel-Plate Media," *Proceedings I.E.E.*, Paper No. 896 R, March 1950 (**97**, Part III, p. 72).

(13) Epstein, D. J.: "Phase Shift of Microwaves in Passage through Parallel Plate Arrays" (Technical Report No. 52, Laboratory for Insulation Research, M.I.T., August 1950).

(14) Lengyel, B. A.: "Reflection and Transmission at the Surface of Metal Plate Arrays," *Journal of Applied Physics*, 1951, **22**, p. 265.

(15) Ruze, J., and Young, M.: "Experimental Determination of Reflection Coefficient of Metal Plate Media," *ibid.*, 1952, **22**, p. 277.

(16) Brady, J. J., Pearson, M. D., and Peoples, S. R.: "Reflection of Microwaves from Metal Plate Media," *ibid.*, 1952, **23**, p. 964.

(17) Weissfloch, A.: "Ein Transformationssatz über Verlustlose vierpole und seine Anwendung auf der experimentelle Untersuchung von Dezimeter- und Zentimeterwellen Schaltungen," *Hochfrequenztechnik*, 1942, **60**, p. 67.

(18) Oliver, M. H.: "Discontinuities in Concentric-Line Impedance Measuring Apparatus," *Proceedings I.E.E.*, Paper No. 884 R, January 1950 (**97**, Part III, p. 29).

(19) Feenberg, E.: "Relation between the Nodal Positions and Standing Wave Ratio in a Complex Transmission System," *Journal of Applied Physics*, 1946, **17**, p. 530.

(20) Cady, Karelitz and Turner: "Radar Scanners and Radomes," Radiation Laboratory Series, **26** (McGraw-Hill Book Co., 1952).

(21) Barlow, H. M., and Cullen, A. L.: "Microwave Measurements" (Constable, 1950).

(11) APPENDIX

(11.1) The Complex Reflection Coefficient for a Symmetrical Slab of a Material

AA will be taken as the reference plane (Fig. 18),

Let n = refractive index of medium b referred to medium a,
l_p = length of b in the direction of propagation,
δ_p = phase delay in $b = (2\pi n l_p)/\lambda_0$,
and λ_0 = wavelength in a.

Then from Reference 20, page 355,

$$\overline{r_{aa}} = R\exp(-j\psi) = \overline{r_{ab}} + \frac{\overline{r_{ba}} \times \overline{t_{ab}} \times \overline{t_{ba}} \times \exp(-j2\delta_p)}{1 - (\overline{r_{ba}})^2 \exp(-j2\delta_p)} \text{ at AA}$$

Put $\overline{r_{ab}} = \rho\exp(-j\phi_{11})$ and $\overline{r_{ba}} = \rho\exp(-j\phi_{22})$

$$\overline{t_{ab}} = \overline{t_{ba}} = \tau\exp(-j\theta)$$

with $\phi_{11} + \phi_{22} - 2\theta = \pm\pi;\ \xi = \phi_{22} + \delta_p$

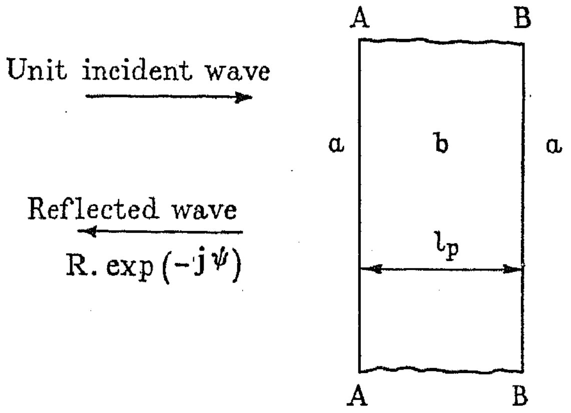


Fig. 18.—Symmetrical slab.

and substitute

$$R\exp(-j\psi) = \rho\exp(-j\phi_{11})\left[1 - \frac{(1-\rho^2)\exp(-j2\xi)}{1-\rho^2\exp(-j2\xi)}\right]$$

which reduces to

$$R\exp(-j\psi) = \frac{2\rho\sin\xi}{\sqrt{[(1-\rho^2)^2 + (2\rho\sin\xi)^2]}}$$

$$\exp\left\{-j\left[\phi_{11} - \arctan\frac{(1-\rho^2)}{(1+\rho^2)\tan\xi}\right]\right\}$$

from which

$$R = \frac{2\rho\sin\xi}{\sqrt{[(1-\rho^2)^2 + (2\rho\sin\xi)^2]}}$$

$$\psi = \phi_{11} - \arctan\left[\frac{(1-\rho^2)}{(1+\rho^2)\tan\xi}\right]$$

and $$\xi = \phi_{22} + \frac{2\pi n l_p}{\lambda_0}$$

(11.2) The Complex Reflection Coefficient for a Composite Slab

AA will be taken as the reference plane (Fig. 19).
a denotes free space.
b denotes the dielectric supporting material of permittivity ϵ.
c denotes plate medium filled with the dielectric material.
n = refractive index of c referred to a.

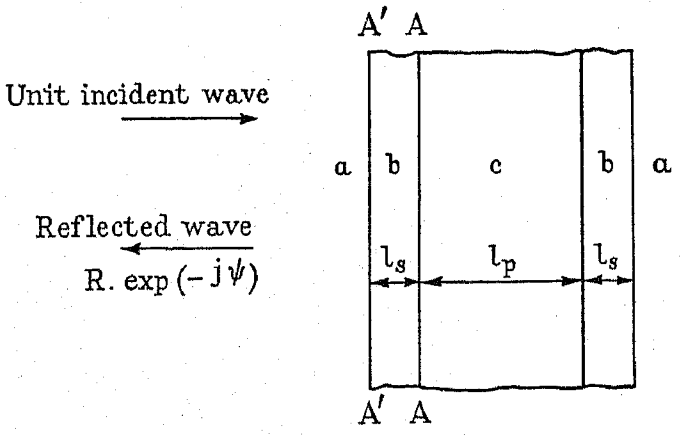


Fig. 19.—Composite slab.

The first step in finding $R\exp(-j\psi)$ is to replace each of the dielectric slabs b by their overall reflection and transmission coefficients referred to the plane AA. (It is assumed that l_s is large enough for any evanescent modes excited at AA to be negligible in amplitude at A′A′.)

P, Φ_{11} and Φ_{22} are found by considering the slabs b.

Put $\overline{r_{ac}} = P \exp(-j\Phi_{11}), \quad \overline{r_{ca}} = P \exp(-j\Phi_{22})$

$$\overline{t_{ac}} = \sqrt{(1 - P^2)} \exp(-j\mu)$$

with $\Phi_{11} + \Phi_{22} - 2\mu = \pm \pi$

and using Section 11.1,

$$R \exp(-j\psi) = \frac{2P \sin \xi}{\sqrt{[(1 - P^2)^2 + (2P \sin \xi)^2]}} \exp\left\{-j\left[\Phi_{11} - \arctan \frac{1 - P^2}{(1 + P^2) \tan \xi}\right]\right\}$$

$$\xi = \delta_p + \Phi_{22}, \quad \delta_p = 2\pi n l_p / \lambda_0$$

The reflection and transmission coefficients at AA are obtained from the Carlson–Heins theory for a plate system completely immersed in dielectric, i.e. use $\lambda_0/\sqrt{\epsilon}$ instead of λ_0.

$$\left.\begin{array}{l}\text{Put } \overline{r_{bc}} = \rho \exp(-j\phi_{11}); \quad \overline{r_{cb}} = \rho \exp(-j\phi_{22}); \\ \overline{t_{bc}} = \tau \exp(-j\theta) = \overline{t_{cb}}; \text{ with } \phi_{11} + \phi_{22} - 2\theta = \pm \pi \end{array}\right\} \text{at AA}$$

At A′A′ the coefficients are simply

$$\overline{r_{ab}} = -\rho_\epsilon = -\frac{\sqrt{(\epsilon)} - 1}{\sqrt{(\epsilon)} + 1}$$

$$\overline{r_{ba}} = \rho_\epsilon = -\overline{r_{ab}}$$

$$\overline{t_{ab}} = \overline{t_{ba}} = \sqrt{(1 - \rho_\epsilon^2)}$$

and $$\delta_s = \frac{2\pi\sqrt{(\epsilon)}l_s}{\lambda_0}$$

$$\overline{r_{ca}} = P \exp(-j\Phi_{22}) = \overline{r_{cb}} + \frac{\overline{t_{cb}} \times \overline{t_{bc}} \times \overline{r_{ba}} \times \exp(-j2\delta_s)}{1 - \overline{r_{ba}} \times \overline{r_{bc}} \times \exp(-j2\delta_s)}$$

at AA (see Reference 20)

Therefore

$$\overline{r_{ca}} = P \exp(-j\Phi_{22}) = \exp(-j\phi_{22}) \left[\frac{\rho(1 + \rho_\epsilon^2) - \rho_\epsilon(1 + \rho^2) \cos x + j\rho_\epsilon(1 - \rho^2) \sin x}{1 - 2\rho_\epsilon\rho \cos x + (\rho\rho_\epsilon)^2}\right]$$

from which

$$P = \left[\frac{\rho^2 + \rho_\epsilon^2 - 2\rho\rho_\epsilon \cos x}{1 + (\rho\rho_\epsilon)^2 - 2\rho\rho_\epsilon \cos x}\right]^{\frac{1}{2}}$$

$$\Phi_{22} = \phi_{22} - \arctan\left[\frac{\rho_\epsilon(1 - \rho^2) \sin x}{\rho(1 + \rho^2) - \rho_\epsilon(1 + \rho^2) \cos x}\right]$$

$$x = 2\delta_s + \phi_{11}$$

And in a similar way

$$\Phi_{11} = -2\pi\left(\frac{2l_s}{\lambda_0}\right) - \arctan\left[\frac{\rho(1 - \rho_\epsilon^2) \sin x}{(\rho_\epsilon(1 + \rho^2) - \rho(1 + \rho_\epsilon^2) \cos x}\right]$$

(11.3) Calculation of the Theoretical Curves

(11.3.1) Extension of Lengyel's Tables.

Lengyel has tabulated[14] as a function of $x = 2a/\lambda_0$, the quantities n, ρ, ϕ_{11} and ϕ_{22} for $x = 1{\cdot}01$ to $x = 1{\cdot}35$. It was found to be necessary to extend this to $x = 1{\cdot}65$ ($\lambda_0 = 8{\cdot}0$ cm). Some of the main values are given in Table 1. The order of accuracy is estimated to be $\pm 0{\cdot}0002$ which was more than adequate for present purposes.

Table 1

EXTENSION OF LENGYEL'S TABLES

$x = 2a/\lambda_s$	n	ρ	$\phi_{11}/2\pi$	$\phi_{22}/2\pi$
1·36	0·6777	0·1921	−0·1505	0·6162
1·40	0·6999	0·1766	−0·1564	0·6248
1·45	0·7241	0·1600	−0·1641	0·6357
1·50	0·7453	0·1459	−0·1723	0·6468
1·55	0·7641	0·1338	−0·1808	0·6582
1·60	0·7806	0·1232	−0·1900	0·6700
1·65	0·7955	0·1139	−0·1998	0·6874

(11.3.2) The Theoretical Curves.

(11.3.2.1) *Refractive Index.*—For the plate system

$$\beta_1^2 = \epsilon\beta^2 - (\pi/a)^2$$

β_1 = propagation constant for the H_{01} mode in the dielectric filled plates.
β = propagation constant for free space.
$= 2\pi/\lambda_0$.
a = plate spacing.
$n = (\beta_1/\beta) = \sqrt{\epsilon}\sqrt{[1 - (1/x)^2]}$.
$x = 2a\sqrt{(\epsilon)}/\lambda_0 = 2a/\lambda_s$.

Thus if n_0 is the value for $\epsilon = 1$, for a given x, n is found from $n_0 = \sqrt{(\epsilon)}n$. The corresponding free space wavelength is $\lambda_0 = 2a\sqrt{(\epsilon)}/x$.

(11.3.2.2) *The Parameters of the Modified Interface* (P, Φ_{11} and ϕ_{22}).—These were best found by choosing a suitable value of x (from which the corresponding free-space wavelength is at once available), obtaining ρ, ϕ_{11} and ϕ_{22} from Lengyel's tables and substituting in the following formulae:

$$P = \left[\frac{\rho^2 + \rho_\epsilon^2 - 2\rho\rho_\epsilon \cos x}{1 + (\rho\rho_\epsilon)^2 - 2\rho\rho_\epsilon \cos x}\right]^{\frac{1}{2}}$$

$$\Phi_{11} = -2\pi\left(\frac{2l_s}{\lambda_0}\right) - \arctan\left[\frac{\rho(1 - \rho^2) \sin x}{\rho_\epsilon(1 + \rho^2) - \rho(1 + \rho_\epsilon^2) \cos x}\right]$$

at plate edges

$$\Phi_{22} = \phi_{22} - \arctan\left[\frac{\rho_\epsilon(1 - \rho^2) \sin x}{\rho(1 + \rho^2) - \rho_\epsilon(1 + \rho^2) \cos x}\right]$$

$$x = 2\left(\frac{2\pi l_s}{\lambda_s}\right) + \phi_{11}.$$

621.372.2 : 621.315.61 : 621.3.011.5

Paper No. 1702 R
Jan. 1955

THE RELATIVE PERMITTIVITY OF TETRAGONAL ARRAYS OF PERFECTLY CONDUCTING THIN DISCS

By JOHN BROWN, M.A., Associate Member, and WILLIS JACKSON, D.Sc., D.Phil., F.R.S., Member.

(*The paper was first received* 28*th January*, *and in revised form* 11*th May*, 1954.)

SUMMARY

The methods available for the calculation of the interaction field of an array of dipoles are discussed with reference to an artificial dielectric formed from a tetragonal array of conducting discs. A modified method is described and used to derive a formula for the relative permittivity, valid when the discs are not too closely spaced. For closely packed discs, the fields due to higher-order multipoles become important and a second formula for the relative permittivity is obtained for this case. Comparison with experimental values already obtained shows that the two formulae together enable the calculation of the relative permittivity over the whole range of spacings likely to be used in practice to be made.

LIST OF PRINCIPAL SYMBOLS

a, b = Lattice spacings defined by Fig. 1.
x, y, z = Cartesian co-ordinates.
E_0 = Magnitude of applied electrostatic field.
m = Magnitude of induced dipole moment.
α = Polarizability of disc.
d = Diameter of disc.
ϵ_0 = Absolute permittivity of free space.
P = Magnitude of polarization vector, i.e. total dipole moment per unit volume.
N = Number of discs per unit volume.
ϵ_r = Relative permittivity of array of discs.
E_1 = Interaction field caused by dipoles.
C = Proportionality constant defined by $E_1 = Cm$.
r, s, t = Integers.
α' = Polarizability per unit length of cylinder formed from from a single row of discs.
n = Refractive index.
B = Susceptance of plane of discs.
Y_0 = Characteristic admittance of free space.
λ_0 = Free-space wavelength.
β = Free-space phase-change coefficient $= 2\pi/\lambda_0$.
$\delta(x)$ = Dirac delta function.
$K_0(x)$ = Modified Bessel function of second kind and zero order.
S = Series defined by eqn. (29).
T_s = Series defined by eqn. (31).

The M.K.S. rationalized system of units is used throughout.

(1) INTRODUCTION

In a recent paper El-Kharadly and Jackson[1] have described the results of a comprehensive investigation into the properties of artificial dielectrics comprising arrays of conducting elements. The particular elements considered included long cylinders and strips, spheres, discs and short cylindrical rods. Experimental values of the relative permittivity were obtained for each of these and in most cases theoretical formulae, in good agreement with these results, were given. An exception was the tetragonal array of discs for which the theory was inadequate except in a limited range of disc diameters and spacings. The theoretical result given by El-Kharadly and Jackson was obtained by the method used by Lorentz and others[2,3,4] in studies of the dielectric properties of non-polar media and assumes that the disturbing action of each disc on a uniform applied electrostatic field can be allowed for if each disc is replaced by a suitably chosen dipole.

The calculation of the relative permittivity of an array of conductors by the Lorentz method is examined in the first half of the paper and an alternative approach, which leads to a new formula, is described. This formula is in reasonable agreement with the experimental results when the spacing b, Fig. 1, is greater than $0\cdot6$ times the spacing a. One of the principal results of the investigation referred to above is that the representation of the conducting obstacles by dipoles becomes inaccurate as the number of elements per unit volume increases and that additional short-range interaction forces then become increasingly important.

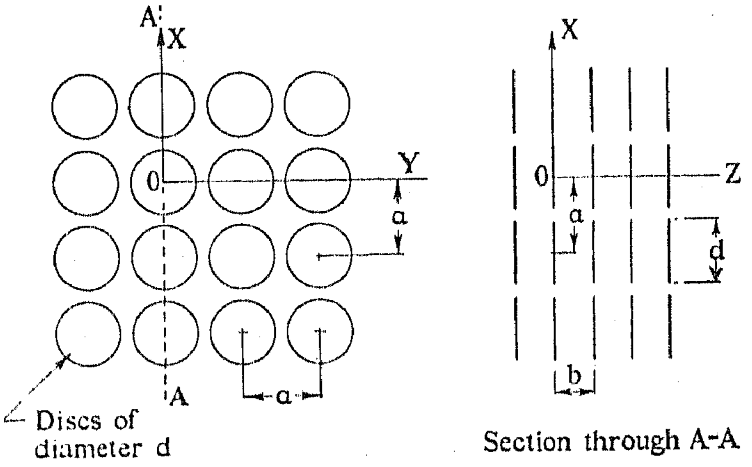


Fig. 1.—Tetragonal array of discs.

In the second part of the paper a method is outlined by which certain of these forces may be allowed for in the calculation of the relative permittivity, leading to theoretical values when the ratio b/a is less than $0\cdot6$.

(2) APPLICATION OF THE LORENTZ THEORY TO ARTIFICIAL DIELECTRICS

When a uniform electric field of strength E_0 volts per metre is applied to the array of Fig. 1 in a direction parallel to the x-axis, a distribution of charge is established on the surface of each disc. The total charge on each disc remains zero but is distributed in such a way that an additional field is created. A first approximation to this extra field is obtained by examining

---

Written contributions on papers published without being read at meetings are invited for consideration with a view to publication.

Mr. Brown is in the Department of Electrical Engineering at Imperial College, University of London.

Dr. Willis Jackson, who was formerly Professor of Electrical Engineering at Imperial College, University of London, is now with the Metropolitan-Vickers Electrical Co., Ltd.

the charge distribution on an isolated disc in a uniform field, and it is found that it is equivalent to a dipole located at the centre of the disc. The dipole is parallel to the applied field, and its moment, m, is proportional to the magnitude of the applied field, i.e.

$$m = \alpha E_0 \quad . \; . \; . \; . \; . \; . \; . \quad (1)$$

The constant of proportionality, α, is called the polarizability, and for a disc of diameter d it has the value $\frac{2}{3}\epsilon_0 d^3$.

Each of the discs in the array of Fig. 1 becomes polarized under the action of the applied field. The polarization, P, is the total dipole moment per unit volume, so that, if there are N discs per unit volume,

$$P = Nm = N\alpha E_0 \quad . \; . \; . \; . \; . \; . \; . \quad (2)$$

For an array in Fig. 1,

$$N = 1/a^2 b \quad . \; . \; . \; . \; . \; . \; . \; . \quad (3)$$

The relative permittivity, ϵ_r, is related to the electric field strength and the polarization by the equation

$$\epsilon_r E_0 = E_0 + P/\epsilon_0 \quad . \; . \; . \; . \; . \; . \; . \quad (4)$$

and substitution for P from eqn. (2) gives

$$\epsilon_r = 1 + N\alpha/\epsilon_0 \quad . \; . \; . \; . \; . \; . \; . \quad (5)$$

There are two limitations on the application of eqn. (5):

(*a*) The assumption that the charge distribution on the surface of each disc corresponds to that of a dipole.

(*b*) The assumption that eqn. (1) applies to discs arranged in an array as well as to isolated discs.

If the first of these assumptions is accepted then the second can be removed by a method which has been investigated by many workers. This method consists simply of calculating the additional field acting on each disc because of the dipoles induced in each of the other discs. Suppose that one of the discs, say the one whose centre is at the origin, is removed, the array being otherwise unaltered. The field at the origin is now $E_0 + E_1$, where E_1 is the interaction field caused by all the dipoles except that at the origin. It is obvious that E_1 is proportional to m, so that

$$E_1 = Cm \quad . \; . \; . \; . \; . \; . \; . \; . \quad (6)$$

where C is a constant of proportionality which has to be determined. If the disc at the origin is now replaced it will be acted on by the total field $E_0 + E_1$, so that eqn. (1) must be replaced by

$$m = \alpha(E_0 + E_1) = \alpha(E_0 + Cm)$$

and if this equation is solved for m there results

$$m = \frac{\alpha E_0}{1 - \alpha C} \quad . \; . \; . \; . \; . \; . \; . \quad (7)$$

The equation for the relative permittivity now becomes

$$\epsilon_r = 1 + \frac{N\alpha}{\epsilon_0(1 - \alpha C)} \quad . \; . \; . \; . \; . \quad (8)$$

The polarizability, α, and the number of elements per unit volume, N, are known for the array of discs and so the relative permittivity can be calculated if a value can be found for the constant C. The calculation of C is far from straightforward and has been the subject of many papers. At first sight, it seems sufficient to calculate the field at the origin due to a dipole at the point (ra, sa, tb), where r, s and t are any integers, and then to sum over all the dipoles excluding that at the origin. This gives the result

$$C = \frac{1}{4\pi\epsilon_0} \sum_{r=-\infty}^{\infty} \sum_{s=-\infty}^{\infty} {\sum_{t=-\infty}^{\infty}}' \frac{2r^2a^2 - s^2a^2 - t^2b^2}{(r^2a^2 + s^2a^2 + t^2b^2)^{5/2}} \quad . \quad (9)$$

the dash signifying the omission of the term for which r, s, t are all zero. Unfortunately this series is only conditionally convergent and different values can be obtained for its sum, depending on the order in which the summations are performed. This difficulty was overcome by Lorentz by the artifice of replacing the assembly of dipoles by a continuous dielectric and then removing a sphere at the origin.[2,3,4] The field within this sphere due to the polarization of the dielectric is $P/3\epsilon_0$, so that, from eqn. (6),

$$C = P/3\epsilon_0 m = N/3\epsilon_0 \quad . \; . \; . \; . \quad (10)$$

Insertion of this value into eqn. (8) gives

$$\epsilon_r = 1 + \frac{N\alpha/\epsilon_0}{1 - N\alpha/3\epsilon_0} \quad . \; . \; . \; . \quad (11)$$

which is known as the Clausius–Mosotti relation. A striking feature of this relation is that it is independent of the ratio b/a. An examination of the conditions under which eqn. (11) is valid shows, however, that it only applies in fact when a and b are equal.[3]

The mathematical equivalent to the foregoing physical approach is to replace the summations, which correspond to the discrete array of dipoles, by integrations, corresponding to a continuous distribution of polarization as in a solid dielectric. It is obvious that this removes any fine detail in the distribution of polarization, and it is inevitable that this process will result in an answer independent of the ratio b/a. The calculation of C by integration in this way was carried out by Rayleigh[5] and the answer obtained is identical with that in eqn. (10).

More recently, Lewin[6] has examined the propagation of electromagnetic waves through a lattice of spheres and has, in the course of this, obtained an equation which is equivalent to eqn. (11). His analysis also involves the approximation of summations by integrations and for the present purpose is effectively the same as that of Rayleigh.

There remains the direct summation of the series, choosing the order of the summations in such a way that the answer obtained agrees with eqn. (10) when a and b are equal. This offers the possibility of investigating the dependence of the factor C on the ratio b/a. Calculations have been carried out in this way, and the values of relative permittivity which are obtained have been compared by El-Kharadly and Jackson with their experimental results. It is obvious from Fig. 13 of their paper that the agreement is far from satisfactory except in the relatively uninteresting region where b/a exceeds 0·8. It is not surprising that a direct summation of the series fails to give results in agreement with experiment, because of the conditional nature of the convergence of the series. It might be possible to obtain more reliable values for C by inserting convergence factors in the series and taking an appropriate limit in the final answer. The mathematical complexity of this approach makes it of doubtful value, at any rate for the investigation of an array of discs.

(3) MODIFIED THEORY

As mentioned in the previous paragraph, the calculation of the relative permittivity of the disc array by summing the series of eqn. (9) gives results which show poor agreement with the measured values when the ratio b/a is less than 0·8. A slightly different approach, which avoids the difficulty of handling a conditionally convergent series, has therefore been tried. This consists simply of the assumption that the principal contribution to the interaction field acting on the disc at the origin comes from those discs which lie in the same plane, i.e. the plane $z = 0$ in Fig. 1. Physically this seems plausible since the separation between the nearest points of adjacent discs is less if they lie in the same plane than if they are in different planes. This is hardly a sufficient justification by itself for adopting this procedure, and a more rigorous analysis, which leads to a similar conclusion, is given in Section 7.1.

If the interaction field is calculated on the assumption that only the discs in the plane $z = 0$ need be considered, the summation in eqn. (9) reduces to the double sum

$$C' = \frac{1}{4\pi\epsilon_0 a^3} \sum_{r=-\infty}^{\infty} {\sum_{s=-\infty}^{\infty}}' \frac{2r^2 - s^2}{(r^2 + s^2)^{5/2}} \quad . \quad . \quad (12)$$

the dash showing the omission of the term for which r and s are both zero. The series in the above equation is absolutely convergent, and there is, in principle, no difficulty in finding its sum. It has not been found possible to obtain an exact answer for the sum but by a method similar to that used by Lewin[7] in studies of waveguide problems, a sufficiently good approximation has been derived. The details of the calculation are given in Section 7.2, where it is shown that the sum has the value 4·521. The approximations used are such that the exact value is less than this and that the difference between the exact and approximate values is not greater than 0·003. From this it follows that

$$C' = 4{\cdot}521/4\pi\epsilon_0 a^3 \quad . \quad . \quad . \quad . \quad . \quad . \quad (13)$$

This expression can be inserted in eqn. (8), and if at the same time the value $2\epsilon_0 d^3/3$ is substituted for the polarizability, the following equation for the relative permittivity results:

$$\epsilon_r = 1 + \frac{a}{b[1{\cdot}5(a/d)^3 - 0{\cdot}360]} \quad . \quad . \quad . \quad (14)$$

(4) THE RELATIVE PERMITTIVITY FOR AN ARRAY OF CLOSELY SPACED DISCS

It is not possible to maintain the assumption that the charge distribution on the surface of each disc corresponds to that of a dipole if the spacing between successive planes of discs (the spacing b in Fig. 1) is smaller than the diameters of the discs. A possible method of allowing for this, theoretically, is to postulate a series of multipoles at the centre of each disc, as described in Section 2.1.3 of Reference 1. The difficulty of formulating this mathematically in the present case is considerable, if not prohibitive. Once again, a physical picture is used to determine what the most important part of the interaction field will be. It is reasonable to suppose that when the spacing b is very small compared with the diameters of the discs, a single row of discs, such as that formed by the discs whose centres lie on the z-axis of Fig. 1, will behave not very differently from a continuous conducting cylinder. This suggests that the array of discs may be regarded as a 2-dimensional lattice of "cylinders," each cylinder consisting of a single row of discs. The calculation of the permittivity of the array of discs can be carried out as follows: first the effective polarizability per unit length of each cylinder, α', is first obtained and then this value of α' is then inserted in the 2-dimensional formula which corresponds to eqn. (8). The appropriate formula for the second step is eqn. (10) of Reference 1, namely

$$\epsilon_r = 1 + \frac{\alpha' N/\epsilon_0}{1 - \alpha' N/2\epsilon_0} \quad . \quad . \quad . \quad . \quad (15)$$

in which N, the number of cylinders per unit area, has the value $1/a^2$ in the present case.

The calculation of the effective polarizability per unit length for the row of discs involves an investigation of the field distribution when the row of discs is placed in a uniform field, of strength E_0 volts per metre. The analysis involved is somewhat lengthy and only the results will be quoted here. It is demonstrated that the assumption that the row of rods behaves like a cylinder is justified, at any rate for small values of the ratio b/d. Further, an approximate expression for the effective polarizability per unit length of the row of rods is

$$\alpha' = \frac{\pi\epsilon_0 d^2}{2}\left[\left(1 - \frac{0{\cdot}441b}{d}\right)^2 + \frac{0{\cdot}058b^3}{d^3}\left(1 + \frac{0{\cdot}441b}{d}\right)\right] . \ (16)$$

The formula is obtained by expressing α' as a series of powers of b/d and retaining powers up to the fourth in the calculations. When b is zero, the value for α' is $\frac{1}{2}\pi\epsilon_0 d^2$, which is identical to that for a continuous conducting cylinder of diameter d.

The relative permittivity for a cuboidal array of discs may be calculated by substituting the above value of α' in eqn. (15). The conditions under which the answer obtained is reasonably accurate are discussed below.

(5) DISCUSSION OF THE RESULTS

Two formulae have been obtained above for the relative permittivity of an array of conducting discs. The first, eqn. (14), is applicable for large values of the ratio b/a, and the second, given by the combination of eqns. (15) and (16), for small values of this ratio. It is very difficult to assess theoretically the range of values of the ratio b/a for which these formulae can be used, but numerical calculation has shown that the difference between them is least when b/a is about 0·6 and that this difference does not exceed 2%. Accordingly, the first formula has been taken as applicable when b/a exceeds 0·6, and the second when b/a is less than 0·6. Numerical values for the relative permittivity, calculated on this basis, are plotted in Fig. 2 against b/a for various values of the ratio d/a. The experimental results obtained by El-Kharadly and Jackson[1] are also shown, and it is

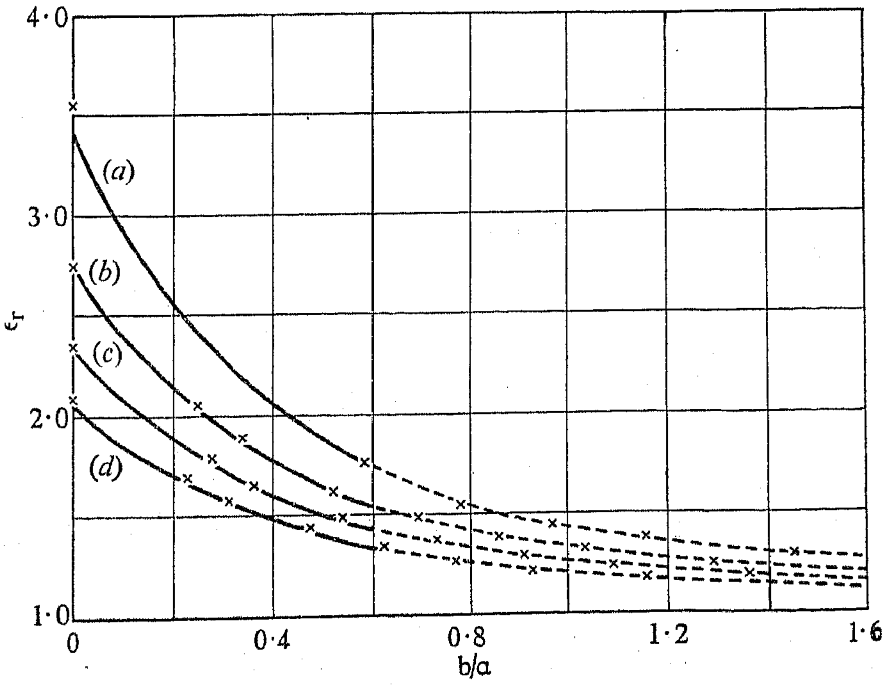


Fig. 2.—Relative permittivity of tetragonal array of discs.

——— Calculated from eqns. (15) and (16).
- - - - Calculated from eqn. (14).
× × × Experimental values from Reference 1.

d/a
(a) 0·833.
(b) 0·770.
(c) 0·714.
(d) 0·667.

seen that the agreement is extremely close over the whole range of b/a.

For the largest value of d/a, only one experimental point—that for the continuous cylinder—lies in the range of b/a from 0 to 0·6. This point differs from the theoretical value by more than the other points, the reason being that the two-dimensional Clausius–Mossoti expression, eqn. (15), ceases to be sufficiently accurate. A modification to this equation allowing for short-range interaction forces between the cylinders gives better agreement and may be used in the present analysis. Since the worst error for a fixed d/a value always arises when b/a is zero, it has not been considered necessary to do this.

Only tetragonal arrays of discs have been considered, since this is the most interesting case in practice, but the methods used can be extended to more general arrays.

(6) REFERENCES

(1) El-Kharadly, M. M. Z., and Jackson, W.: "The Properties of Artificial Dielectrics comprising Arrays of Conducting Elements," *Proceedings I.E.E.*, Paper No. 1472 R, July 1953 (**100**, Part III, p. 199).
(2) Debye, P.: "Polar Molecules" (Dove Publications, 1945).
(3) Fröhlich, H.: "Theory of Dielectrics" (Oxford University Press, 1949).
(4) Böttcher, C. J. F.: "Theory of Electric Polarisation" (Elsevier Publishing Company).
(5) Rayleigh, J. W. S.: "On the Influence of Obstacles arranged in Rectangular Order upon the Properties of a Medium," *Philosophical Magazine*, 1892, **34**, p. 481.
(6) Lewin, L.: "Electrical Constants of Material loaded with Spherical Particles," *Journal I.E.E.*, 1947, **94**, Part III, p. 65.
(7) Lewin, L.: "Advanced Theory of Waveguides" (Iliffe Press, 1951).
(8) Brown, J.: "The Design of Metallic Delay Dielectrics," *Proceedings I.E.E.*, Paper No. 915 R, January 1950 (**97**, Part III, p. 45).
(9) Cohn, S. B.: "Electrolytic-Tank Measurements for Microwave Delay Lens Media," *Journal of Applied Physics*, 1950, **21**, p. 674.
(10) Stratton, J. A.: "Electromagnetic Theory" (McGraw-Hill, 1941).
(11) Titchmarsh, E. C.: "The Fourier Integral" (Oxford University Press, 1937).
(12) Jahnke, E., and Emde, F.: "Table of Functions" (Dover Publications, 1943).

(7) APPENDICES

(7.1) Derivation of the Interaction Factor from Transmission-Line Analysis

It has been shown[8] that the propagation of an electromagnetic wave of free-space wavelength λ_0 in the direction of the z-axis of Fig. 1 can be represented by the loaded transmission line of Fig. 3. Each shunt susceptance, jB, corresponds to a plane of discs, and these susceptances are independent, i.e. there is no reactive coupling between adjacent ones, if the spacing b exceeds $0 \cdot 75a$.[9] A general formula for the refractive index of the array represented by the circuit of Fig. 3 has been derived, and when the wavelength is much larger than the spacing b, it reduces to

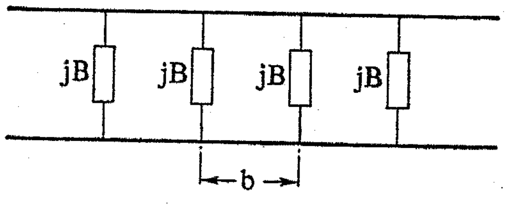


Fig. 3.—Equivalent circuit for array of discs. The transmission line has a phase-change coefficient β_0 and a characteristic admittance Y_0.

$$n^2 = 1 + B\lambda_0/2\pi b Y_0 \quad . \quad . \quad . \quad . \quad . \quad (17)$$

Y_0 being the characteristic admittance of the transmission line. Kock has pointed out that the permeability of an array of discs, under the conditions being discussed, is the same as that of free space and so n^2 is equal to the relative permittivity, ϵ_r.

The value of the shunt susceptance may be calculated in terms of the electrostatic polarizability of the individual discs. It is sufficient to consider a single plane of discs, say that given by $z = 0$. If a plane wave polarized so that its electric field is parallel to the x-axis, and travelling in the direction of the z-axis, is incident upon this plane, then each disc will act as if there were an oscillating dipole at its centre. The axis of the dipole is parallel to the incident electric field. The magnitude of the dipole moment, m, may be calculated in a way similar to that used in the electrostatic analysis given in Sections 2 and 3, i.e.

$$m = \alpha(E_0 + E_1) \quad . \quad . \quad . \quad . \quad . \quad (18)$$

where α is the polarizability.

E_0 is the amplitude of the incident field at the origin.
E_1 is the amplitude of the field at the origin, caused by the presence of the dipoles.

Electrostatic formulae may be used in calculating the interaction field E_1, since the principal contribution to it arises from the dipoles, whose distance from the origin is of the order of a and, therefore, small compared with the wavelength. For the same reason, it is permissible to use the electrostatic value for the polarizability. The interaction field may be written

$$E_1 = C'm \quad . \quad . \quad . \quad . \quad . \quad . \quad (19)$$

where $$C' = \frac{1}{4\pi\epsilon_0 a^3} \sum_{r,s=-\infty}^{\infty}{}' \frac{2r^2 - s^2}{(r^2 + s^2)^{5/2}} \quad . \quad . \quad . \quad (20)$$

The interaction field may be eliminated from eqns. (18) and (19), giving

$$m = \frac{\alpha E_0}{1 - \alpha C'} \quad . \quad . \quad . \quad . \quad . \quad . \quad (21)$$

It now remains to determine the relation between the susceptance, B, and the dipole moment, m. The oscillating dipoles set up a reflection field of amplitude ρE_0, where ρ is the reflection coefficient, and an evanescent field confined to the neighbourhood of the plane $z = 0$. The complete field must satisfy the inhomogeneous wave equation[10]

$$\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} + \beta^2\right)E_x = -\frac{\beta^2 P_x}{\epsilon_0} \quad . \quad . \quad (22)$$

where β (which is equal to $2\pi/\lambda_0$) is the free-space phase-change coefficient. The quantity P_x on the right-hand side is the distribution of dipole moment, directed parallel to the x-axis, and is zero everywhere except at the centre of each disc where there is concentrated a dipole of moment m. This is expressed mathematically by the relation

$$P_x = m\delta(z) \sum_{r,s=-\infty}^{\infty} \delta(x - ra)\delta(y - sa) \quad . \quad . \quad (23)$$

where $\delta(x)$, etc., are Dirac delta functions.

The principal part of the electromagnetic field is

$$\left.\begin{aligned} E_x &= E_0 \exp(-j\beta z) + \rho E_0 \exp(j\beta z) & \text{if } z \leqslant 0 \\ &= E_0(1 + \rho)\exp(-j\beta z) & \text{if } z \geqslant 0 \end{aligned}\right\} \quad . \quad (24)$$

and the evanescent part has the period a in both x and y because of the same periodicity in P_x. Integration of both sides of eqn. (22) with respect to x and y for the range $-a/2$ to $a/2$ in each case therefore gives

$$\left.\begin{aligned} a^2E_0\left(\frac{\partial^2}{\partial z^2}+\beta^2\right)[\exp(-j\beta z)+\rho\exp(j\beta z)] &= -\frac{\beta^2 m\delta(z)}{\epsilon_0} \\ a^2E_0(1+\rho)\left(\frac{\partial^2}{\partial z^2}+\beta^2\right)\exp(-j\beta z) &= -\frac{\beta^2 m\delta(z)}{\epsilon_0} \end{aligned}\right\} \quad (25)$$

A further integration over a vanishingly small interval of z containing the origin is now performed. The only contribution on the left-hand side arises from the discontinuity in the derivative with respect to z at the origin, and so

$$2j\beta\rho a^2E_0 = \beta^2 m/\epsilon_0$$

i.e. $$\rho = -j\beta m/2\epsilon_0 a^2E_0 = -j\beta\alpha/2\epsilon_0 a^2(1-\alpha C') \quad . \quad (26)$$

The reflection coefficient is related to the susceptance by the ordinary transmission line formula, and since in the present case B/Y_0 is much less than unity, the approximation

$$B/Y_0 = 2j\rho \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (27)$$

may be used. Substitution of the results of eqns. (26) and (27) in eqn. (17) gives

$$\epsilon_r = 1 + \frac{N\alpha}{\epsilon_0(1-\alpha C')} \quad . \quad . \quad . \quad . \quad (28)$$

since $1/a^2b$ is equal to N the number of elements per unit volume. The above formula is valid if there is no reactive coupling between successive susceptances and, therefore, certainly applies when b exceeds $0{\cdot}75a$.

(7.2) Evaluation of the Double Series

The series to be evaluated is

$$S = \sum_{r,s=-\infty}^{\infty}{}' \frac{2r^2-s^2}{(r^2+s^2)^{5/2}} \quad . \quad . \quad . \quad . \quad (29)$$

the dash signifying the exclusion of the term for which r and s are both zero. This series is absolutely convergent and may be summed in any order. It may therefore be written

$$S = \sum_{r=-\infty}^{\infty} 2/|r|^3 - 2\sum_{S=1}^{\infty} T_s \quad . \quad . \quad . \quad . \quad (30)$$

the terms for which s is zero being summed separately. The fact that the value of any term is unaltered if the sign of s is changed, is used to restrict the range over which the summation with respect to s has to be performed. The partial sums T_s are defined by the equation

$$T_s = -\sum_{r=-\infty}^{\infty}\frac{2r^2-s^2}{(r^2+s^2)^{5/2}} \quad . \quad . \quad . \quad . \quad (31)$$

This series can be summed approximately by using Poisson's theorem[11] which states that if $F(u)$ is the Fourier cosine transform of $f(x)$, i.e.

$$F(u) = \sqrt{\left(\frac{2}{\pi}\right)}\int_0^{\infty} f(x)\cos ux\,dx \quad . \quad . \quad . \quad (32)$$

then $$\sum_{r=-\infty}^{\infty} f(r) = \sqrt{(2\pi)}\left[F(0) - 2\sum_{n=1}^{\infty}F(2n\pi)\right] \quad . \quad (33)$$

In the present case $f(x)$ is taken as $(2x^2-s^2)/(x^2+s^2)^{5/2}$ and $F(u)$ is found to be $-\sqrt{(2/\pi)}u^2K_0(us)$, where $K_0(us)$ is a modified Bessel function of the second kind. The value for $F(0)$ must be interpreted as the limit of $F(u)$ as u tends to zero and since $K_0(us)$ is of the order of $1/u$ for small values of u, it follows that $F(0)$ is zero. Hence from eqns. (31) and (33)

$$T_s = 16\pi^2\sum_{n=1}^{\infty} n^2K_0(2n\pi s) \quad . \quad . \quad . \quad . \quad (34)$$

The series on the right of the above equation is very rapidly convergent, and although an exact expression for the sum has not been obtained, a sufficiently good approximate value is easily calculated. The modified Bessel function can be expressed as an infinite integral, namely

$$K_0(2\pi ns) = \int_0^{\infty}\exp(-2\pi ns\cosh\phi)d\phi \quad . \quad . \quad (35)$$

This integral is substituted in the series, and the order in which the integration and the summation are performed is reversed, so that

$$T_s = 16\pi^2\int_0^{\infty}\left[\sum_{n=1}^{\infty} n^2\exp(-2\pi ns\cosh\phi)\right]d\phi \quad . \quad (36)$$

The series under the integral sign is a standard one, whose sum is known and so

$$T_s = 16\pi^2\int_0^{\infty}[E(1+E)/(1-E)^3]d\phi \quad . \quad . \quad (37)$$

where $$E = \exp(-2\pi s\cosh\phi) \quad . \quad . \quad . \quad . \quad (38)$$

Limits for the value of T_s can be obtained from this integral. Since $\cosh\phi$ exceeds unity for real values of ϕ, E is never greater than E_0, which equals $\exp(-2\pi s)$ and is small compared with unity. A first approximation for T_s is therefore obtained by neglecting E compared with unity, giving

$$\begin{aligned} T_s &= 16\pi^2\int E d\phi \\ &= 16\pi^2K_0(2\pi s) \quad . \quad . \quad . \quad . \quad . \quad (39) \end{aligned}$$

i.e. just the first term of the series. Since all terms of the series are positive, this must be a lower limit to the sum. An upper limit is obtained as follows:

$$1+E \leqslant 1+E_0 : 1-E \geqslant 1-E_0$$

Hence,

$$\int_0^{\infty}[E(1+E)/(1-E)^3]d\phi \leqslant \int_0^{\infty}[E(1+E_0)/(1-E_0)^3]d\phi$$

$$\leqslant [(1+E_0)/(1-E_0)^3]\int_0^{\infty}E d\phi = [(1+E_0)/(1-E_0)^3]K_0(2\pi s)$$

An upper limit for the sum of the series is therefore

$$T_s' = 16\pi^2(1+E_0)K_0(2\pi s)/(1-E_0)^3 \quad . \quad . \quad . \quad (40)$$

The error which arises from replacing the series by its first term may be estimated from eqns. (39) and (40). The worst case is that for s equal to unity, and numerical calculation shows that the error is then not worse than $0{\cdot}75\%$. It suffices then to take as the sum T_s the value given by eqn. (39).

Substitution of this in eqn. (30) gives

$$S = 4\sum_{r=1}^{\infty} 1/r^3 - 32\pi^2 \sum_{s=1}^{\infty} K_0(2\pi s) \quad . \quad . \quad . \quad (41)$$

The first series is a particular case of the Riemann zeta function, whose value is known. The second series is very similar to that for T_s and it may be summed approximately by the method which has just been described. The error in replacing the series by its first term is not greater than 0·2%. The sum S is therefore given by $4\sum_{r=1}^{\infty} 1/r^3 - 32\pi^2 K_0(2\pi)$, and the error in replacing the double series in eqn. (30) by a single term does not exceed 1%. Substitution of numerical values gives

$$4\sum_{r=1}^{\infty} 1/r^3 = 4\cdot804 \qquad (\text{Ref. } 12)$$

$$2\sum_{s=1}^{\infty} T_s = 32\pi^2 K_0(2\pi) \qquad (-0,\ +1\%)$$

$$= 0\cdot283 \qquad (-0,\ +0\cdot003)$$

$$S = 4\cdot521 \qquad (+0,\ -0\cdot003)$$

The error in the final expression for S is less than 0·1%, which is more than adequate for a comparison with the experimental results.

DISCUSSION ON
"THE EFFECT OF IRRADIATION ON THE CALIBRATION OF 2CM-DIAMETER SPHERE-GAPS"*

Messrs. E. A. Smith and **D. E. S. Hill** (*communicated*): We have recently had cause to use a 2cm sphere-gap for the measurement of the peak value of 1/5 negative impulse voltage waves up to 30kV. Relying solely upon the irradiation derived from the gaps of the impulse generator, which in the case of larger-diameter spheres had always been found adequate, very variable results were obtained. It was therefore decided to use a more intense irradiation, and a commercial ultra-violet lamp which was readily available was employed.

The 2cm-diameter steel spheres were positioned 40cm from the quartz tube of the lamp and screened from external irradiation. The calibration (50% values) was then carried out by means of a single-stage impulse generator, the efficiency of which was calculated and then checked oscillographically. The impulse-generator charging voltage was measured to an accuracy of 1% by a microammeter connected in series with a high-stability resistor.

The results of the calibration are compared in Table A with the 50% impulse breakdown voltages (1·0mg radium) given in Table 2 of the paper.

It is of interest to note that for gap lengths of from 0·6 to 0·8cm the impulse breakdown voltages with ultra-violet irradiation are within 1% of those given for irradiation with 1·0mg radium, and this difference could be attributed to experimental error. For gap lengths of from 0·2 to 0·5cm the impulse breakdown voltages with ultra-violet irradiation are somewhat higher—and for gap lengths below 0·2cm somewhat lower—than those with 1·0mg radium irradiation.

Dr. D. R. Hardy and **Mr. T. E. Broadbent** (*in reply*): It is not considered good practice to irradiate sphere gaps used for voltage measurement with the ultra-violet radiations obtained from an impulse-generator spark-gap. The breakdown voltage of a given spark-gap is known to depend on the intensity and spectral distribution of the ultra-violet radiations, which are not easily controlled unless a lamp of a type apparently used by Messrs. Smith and Hill is employed. Contrary to their experience, we consider it more usual to find such a lamp less effective as a source of irradiation than impulse-generator spark-gaps, although in their case the spark may have been weak and the distances large.

Table A

IMPULSE-VOLTAGE CALIBRATION OF 2CM-DIAMETER SPHERE-GAPS

Gap length	50% impulse breakdown voltage	
	1·0mm radium irradiation (Hardy and Broadbent)	Ultra-violet irradiation
cm	kV	kV
0·13	6·7*	6·3
0·15	7·1	7·0
0·2	8·5	8·9
0·3	11·6	12·0
0·4	14·5	15·0
0·5	17·5	17·8
0·6	20·4	20·6
0·7	23·2	23·2
0·8	26·0	26·2

(Corrected to 20°C and 760mm Hg.)

* By interpolation.

Experiments made in the laboratories with which we are associated show results which more or less agree with those of Messrs. Smith and Hill for an irradiation which produces a total gap current of about 3×10^{-14}amp. The fraction of this current which is effective in initiating breakdown has not yet been determined. It may be of interest to note that variations in breakdown voltage in excess of 10% were obtained over a wide range of gap lengths as the total photo-current was increased from 5×10^{-15}amp to about 10^{-12}amp.

The fact that the results given by Messrs. Smith and Hill agree in part with those of Table 2 is, perhaps, a coincidence, and had more or less intense irradiation been employed the agreement might not have been so good. Great care should be exercised when comparing values of breakdown voltage for gaps of 0·2cm and less when ultra-violet irradiation is used, because of the reduction in the value of photo-current due to the small angle of incidence with the surface of the sphere.

---

* Paper by D. R. HARDY and T. E. BROADBENT (see 1954, **101**, Part II, p. 438).

621.396.11.029.5/6

Paper No. 1747 R
Jan. 1955

SOME FEATURES OF V.H.F. TROPOSPHERIC PROPAGATION

By M. W. GOUGH, M.A., Associate Member.

(*The paper was first received 7th April, and in revised form 4th August*, 1954.)

SUMMARY

Widespread v.h.f. continuous signal-strength measurements made in tropical and Mediterranean regions over the last four years have exemplified and thrown into relief many of the well-known mechanisms of tropospheric propagation. Although these mechanisms are broadly familiar to many engineers and physicists, the practical magnitudes, as exemplified throughout the paper, are perhaps less widely appreciated despite their great importance to the communication engineer.

It is the purpose of the paper to present and analyse new v.h.f. data from the sources indicated, so as to show (*a*) that observed v.h.f. signal patterns are consistent with tropospheric structures known commonly to occur, and (*b*) that sufficient information is now available to enable the communication engineer to forecast to a useful accuracy the statistical behaviour of projected v.h.f. radio paths in certain parts of the world.

(1) INTRODUCTION

The influence of the troposphere on the propagation of very short radio waves has been recognized for a long time. As early as 1933 Schelleng, Burrows and Ferrell[1] introduced the conception of the effective earth radius to represent the influence of the atmospheric refractive-index gradient. Later, theoretical workers, including Eckersley and Millington,[2] Domb and Pryce,[3] described modifications to their parameters which adapted their formulae to include the effect of refraction. It was not until 1938, however, that Englund, Crawford and Mumford[4] correlated measured v.h.f. signal strengths with atmospheric refractive-index gradients computed from meteorological data, and at the same time successfully explained long-range v.h.f. propagation in terms of reflections from an elevated atmospheric layer. The greatly expanded applications of very high frequencies during the war revealed anomalous features of propagation which gave a great impetus to radio-meteorology. The phenomena of trapping and scattering, for example, were investigated theoretically by Booker[5,6] and Megaw,[7] who obtained results which were consistent both with controlled tests and a mass of operational radar observations.

Meanwhile the exacting demands of the radio-meteorologist promoted more detailed meteorological investigations of atmospheric stratification, particularly near the ground. Measurements made in the Mediterranean in 1922 by Sir Nelson Johnson were found by Appleton[8] to imply an unexpectedly steep negative refractive-index gradient near the sea surface, thus explaining the trapping of centimetric waves observed in that region during the war. Widespread meteorological soundings collated after the war by Booker, Sheppard,[9] Durst[10] and Saxton,[11] coupled with the results of the 1927 American expedition on the S.S. *Meteor*,[12] have provided much useful material for the radio-meteorologist.

In an analysis of long-range measurements on British v.h.f. stations, Saxton[11] has correlated trends in received signal-strength with average refractive-index gradients and the occurrence of elevated layers. Owing to the difficulty of obtaining adequate meteorological information for the great volume of atmosphere involved in v.h.f. propagation, attempts by many workers at detailed correlation have met with only limited success. The advent of the cavity refractometer,[13] with its ability to measure minute and rapid deviations in atmospheric refractive index, promises increasing success in this field.

As our knowledge of atmospheric fine-structure becomes increasingly comprehensive and topical, the incalculable components of signal variations will, in the absence of mathematical and computational limitations, become progressively reduced until, in the limit, a perfectly random residue[14] only is left. While, from the scientific aspect, the analysis of signal behaviour in the finest detail compatible with available techniques is often of great value, the communication engineer will continue, rightly, to predict the performance of his circuits on the best available statistics. He will, however, rely on the radio-meteorologist for the solution of special problems leading to improvements in the efficiency of radiocommunication. It is the purpose of the present paper to show the ultimate value to the engineer of both these methods of approach as applied to v.h.f. communication problems.

The paper is based on about 15 000 hours of v.h.f. signal-strength records of controlled low-power v.h.f. transmissions made in many parts of the world during the last four years. More than 100 paths, mostly in the tropics, have been surveyed, and a great variety of topographical and tropospheric conditions has provided valuable material for the study of v.h.f. propagation.

The information extracted from these records is of two kinds. First, there is statistical information useful to the communication engineer, e.g. the range of fading likely to be encountered over a radio path in a specific region. Secondly, there is the more fundamental information on tropospheric propagation mechanisms revealed by a detailed analysis of small sections of record. This kind of understanding is essential if improvements in v.h.f. communication are to continue.

Physical aspects of v.h.f. tropospheric propagation are treated in Section 2, where recorded v.h.f. signal patterns are analysed with reference to idealized conditions approximating to those commonly observed in meteorology. Variations in effective earth-radius factor are deduced, and it has been found possible in some cases to separate the effects of elevated reflecting layers, while at the same time deducing their approximate structure. No suitably located meteorological measurements were available for checking these radio-deduced tropospheric reconstructions, but they are all plausible. While the results are inevitably approximate, they demonstrate that the frequently observed large departures from "standard" propagation conditions are quite compatible with the tropospheric structures known to meteorology.

While the radio-meteorological information in Section 2 is by no means complete, it is hoped that it will guide communication engineers in forecasting the general behaviour of projected routes in tropical and Mediterranean regions for which little radio and tropospheric information was hitherto available.

Section 3 surveys the broad features of the v.h.f. signal records from an engineering viewpoint. Typical signal patterns are discussed, and a connection is established between the median attenuation of a path and its propensity to fading under specific climatic conditions. Charts are included giving very approximately the probable fading range of tropical v.h.f. paths in terms

---

Written contributions on papers published without being read at meetings are invited for consideration with a view to publication.
Mr. Gough is with Marconi's Wireless Telegraph Co. Ltd.

of their median path-attenuation, and it is shown that sea paths have a greater tendency to fading than land paths of equal median attenuation. From these charts, in conjunction with knowledge of the average effective earth-radius factor for the region, it is possible to forecast approximately the probable fading range of paths whose geometry is sufficiently simple to allow approximate computation of the path attenuation *in vacuo*.

(1.1) Instrumental Details, Conventions Adopted and Definitions of Terms

To assist in the appraisal of the results, a very brief description follows of the v.h.f. signal-strength measuring equipment used in all but three of the tests discussed.

A crystal-controlled c.w. transmitter gives a continuously monitored output of 2–3 watts into a horizontal Yagi aerial having a gain of 8 dB relative to a half-wave dipole. At the receiving terminal a similar aerial system feeds a crystal-controlled superheterodyne receiver of 70-ohm input impedance, giving an output current approximately proportional to the input signal expressed in decibels above one microvolt. An associated recording microammeter caters for input signals ranging from 0–70 dB/μV. Signals can be sampled at 2- or 1-min intervals and are recorded in the form of a row of dots on a slowly moving paper chart controlled by a clock. The time between sampling is occupied in recording the receiver gain while injecting a known signal level from an associated stabilized signal-generator. The virtually continuous monitoring of receiver gain and transmitter power ensures that the signal recordings, after applying the requisite corrections, reflect true variations in path behaviour. The equipment comprises two types covering the respective bands 80–100 Mc/s and 150–180 Mc/s approximately.

For the Lake Victoria tests (Section 2.2.4) prototype equipment giving a transmitted power of about 8 watts was used. It utilized the same principles as the later equipment described above. The Israel–Cyprus test (Section 2.2.1) utilized a 200-watt c.w. transmitter on 77·47 Mc/s, in conjunction with standard receiving equipment described above. The simultaneous 174-Mc/s test utilized standard equipment at both terminals.

In discussing the records produced by this equipment the interest is often centred on signal variations rather than on absolute values, and therefore little mention is made of instrumental details such as transmitted power, feeder losses and aerial gains. However, where concise engineering information is of value, these experimental constants are conveniently merged with the received signal level in the path attenuation between half-wave dipoles (usually abbreviated to path attenuation).

Imagining idealized loss-free half-wave dipoles terminating the transmission path, path attenuation is defined as

$$10 \log_{10} P/p \text{ decibels}$$

where P is the power to the transmitting dipole and p is the power to a matched receiver fed by the receiving dipole.

When required, the path attenuation is derivable from a measured level S decibels above one microvolt by subtracting the measured figure from a constant which depends on the receiver input impedance and the effective radiated power, where this power includes the effect of aerial gains and feeder losses at both terminals. Table 1 gives the approximate conversion factors for three relevant instrumental conditions.

Table 1

Instrumental conditions	Path attenuation
	dB
Standard equipment used in all but three tests ..	$150 - S$
Israel–Cyprus test on 77·47 Mc/s using high-power transmitter	$175 - S$
Lake Victoria tests using prototype equipment:	
77 Mc/s	$160 - S$
169 Mc/s	$154 - S$

Discussion of tropospheric propagation involves frequent use of the radio-meteorological terms defined below:

n = Atmospheric refractive-index at a height h above the earth.

$\dfrac{dn}{dh}$ = Refractive-index gradient.

R_0 = True earth radius (3 960 miles, or 6 370 km).

$$R_m = \text{Effective or modified earth-radius} = \frac{1}{\dfrac{dn}{dh} + \dfrac{1}{R_0}}$$

$$M = \text{Atmospheric refractive modulus at a height } h \text{ above the earth} = 10^6\left(n - 1 + \frac{h}{R_0}\right)$$

M_0 = value of M at ground level.

$$\frac{dM}{dh} = \text{Refractive modulus gradient or } M\text{-gradient} = 10^6\left(\frac{dn}{dh} + \frac{1}{R_0}\right) = \frac{10^6}{R_m}$$

$$k = \text{Effective earth-radius factor} = \frac{R_m}{R_0} = \frac{10^6}{R_0 \dfrac{dM}{dh}}$$

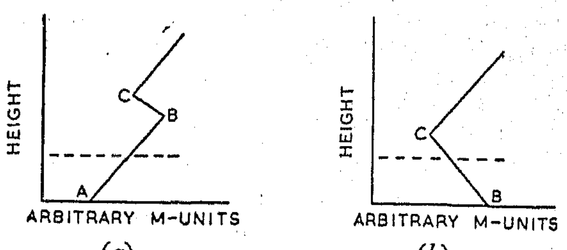
As we are concerned with changes of n and M rather than their absolute values, we shall be dealing throughout mainly with the quantities dn/dh, dM/dh and k, and a brief mention of the probable ranges of these variables will not be out of place. As is usual we shall specify dn/dh and dM/dh in the metric system, but the British system is retained for all heights and distances.

Meteorological measurements show, and radio measurements broadly confirm, that normally we shall be concerned with values of dn/dh ranging from about zero (indicative of marked sub-refraction) to about -3×10^6 per metre, indicative of an intense duct. This implies M-gradients ranging from about 0·16 to $-2 \cdot 8$ M-units/metre, and values of k ranging from 1 through infinity to about $-0 \cdot 055$. It should be noted that, by definition, a duct implies a negative M-gradient, from which it follows that dn/dh in a duct must exceed $-0 \cdot 157 \times 10^6$ per metre.

(2) SOME PHYSICAL ASPECTS OF V.H.F. PROPAGATION

(2.1) Idealized Mechanisms of Tropospheric Propagation

The M-profiles of Fig. 1 represent in idealized form the basic tropospheric conditions governing v.h.f. propagation. Con-



(*a*) (*b*)

Fig. 1.—Basic idealized M-profiles.
— — — — Heights of terminals.

dition (*a*) is dominant in temperate regions, while condition (*b*) will often be the controlling factor over desert terrain and in some tropical and sub-tropical regions, both coastal and inland. Propagation within these types of troposphere has been investigated theoretically and experimentally by many workers,[4,5,8,11] and it will suffice here to describe briefly the application of these established methods to the analysis of v.h.f. signal-strength records summarized in the paper.

Considering first Fig. 1(*a*), the radio terminals lie within a surface stratum AB whose *M*-gradient is approximately constant with height but changes slowly with time. Fundamentally the received signal is controlled by the path geometry and this *M*-gradient, but relatively rapid interference effects will, in general, be superimposed on this basic signal trend by partial reflection at the elevated discontinuity BC. When BC is of sufficient magnitude, its presence is manifest by fading. By smoothing out such fades on a v.h.f. signal-strength record, we are left with a signal trend which is a fair reflection of the *M*-gradient, and implicitly of the effective earth-radius factor *k*, appropriate to the stratum AB. Fig. 2 shows an example of the smoothing process applied to a section of record received over a 200-mile non-optical path between Israel and Cyprus. In this case, simultaneous recording on two frequencies enabled the influence of elevated reflecting layers to be discounted with considerable accuracy, since some uncertainties in the process could be eliminated by ensuring good correlation between the two smoothed signal-patterns.

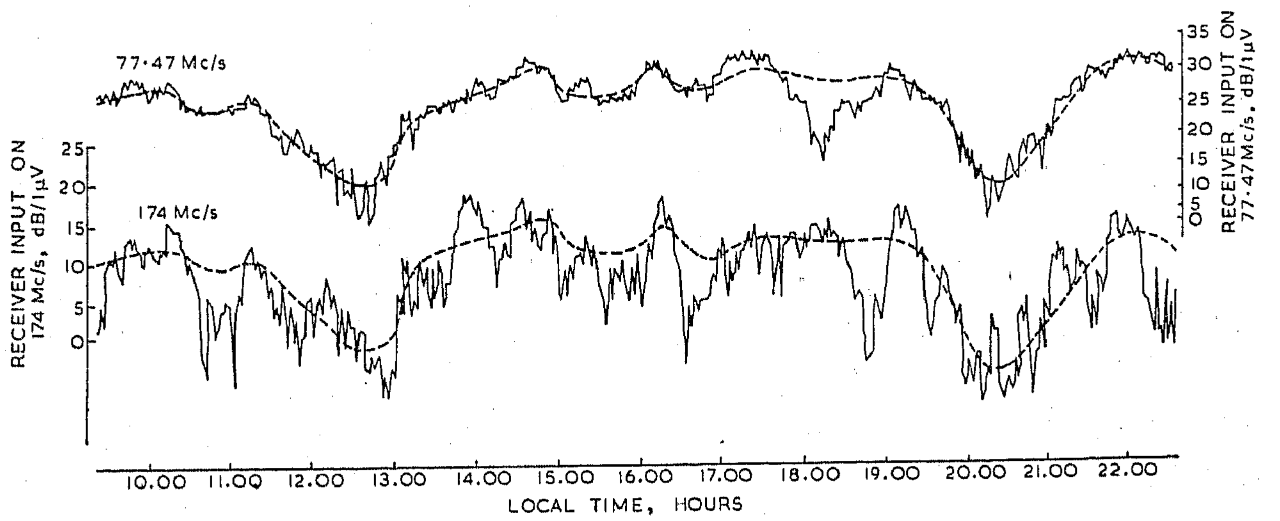


Fig. 2.—Correlated simultaneous signal records.
— — — — Smoothed curves.

Knowledge of the most probable values and range of *k*, averaged over the lower regions of the troposphere, is of considerable importance in planning v.h.f. systems. This information can be extracted from the smoothed signal-strength records by applying them to a curve of theoretical signal-strength calculated in terms of *k* for the path in question. Such curves, calculated for convenience in terms of path attenuation rather than signal strength, are shown in Fig. 3, plotted for the above-mentioned Israel–Cyprus path for the two test frequencies.

The depth of fading over a v.h.f. path, which is, of course, of paramount interest to the communication engineer, is controlled by the relative amplitudes of the "orthodox" signal governed by *k* and the signal reflected from the elevated layer BC [Fig. 1(*a*)]. Various workers[15,16] have idealized layers of this type in mathematically tractable ways for computing the modulus and phase of the reflection coefficient of the layer. The communication engineer is mainly concerned with the modulus, because the phase, being extremely sensitive to the layer characteristics, is for practical purposes indeterminate. Thus, while the exact time of occurrence of fades will always be unpredictable, an approximate knowledge of the modulus of the coefficient will help to forecast general fading behaviour. To indicate the magnitudes likely to be involved in v.h.f. propagation, Fig. 4 shows curves of tropospheric reflection coefficient $|\rho|$ versus the ray-grazing angle θ for suitably idealized layers 36 wavelengths thick, for probable values of Δn, the total change of refractive index across the layer. The curves are derived from charts of tropospheric reflection coefficients due to Millington.[16] The most significant feature of the curves is the onset of total internal reflection for grazing angles below a critical value. These angles are often approached in practice, and the consequent strong-layer reflections can produce deep fading. The variation of layer reflection coefficients with distance, and their connection with fading statistics, are discussed in Section 3.4.

The lack of adequate meteorological measurements along the paths subjected to radio tests has prompted attempts to reconstruct from the radio results the approximate state of the tropo-

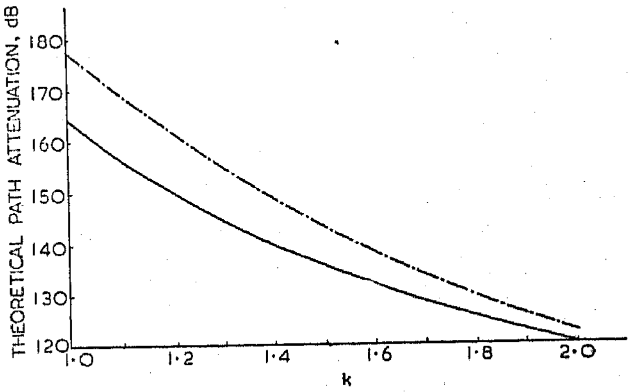


Fig. 3.—Israel–Cyprus path: theoretical variation of path attenuation with *k* for the two test frequencies.
—— 77·47Mc/s.
—·— 174Mc/s.

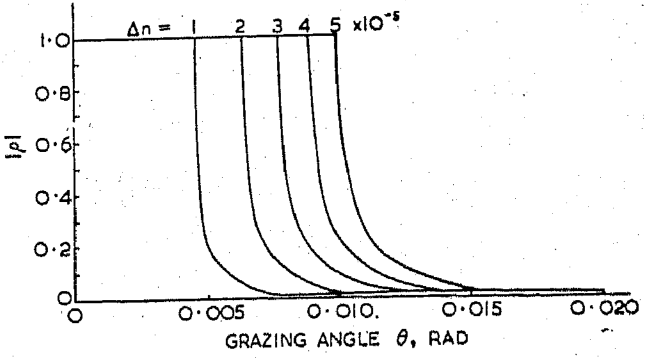


Fig. 4.—Theoretical variation of reflection coefficient with grazing angle for elevated layers 36λ thick.

sphere during the tests. Measurement of the depth of selected fades, and the signal strength immediately before or after the fade, has enabled some tropospheric reflection coefficients to be estimated (Section 2.3), which in turn have in some cases yielded information on the probable heights of layers responsible for the reflections (Section 2.4).

The M-curve of Fig. 1(b), where the terminals lie within a surface duct, will now be considered. Strictly, propagation within the layer BC should be treated by the mode theory of Booker,[17] applying appropriate boundary conditions, but for the relatively short ranges involved in the present paper, a two-ray theory taking no cognizance of the boundary C gives a fairly satisfactory explanation of the observed signal patterns. By considering propagation over an effectively concave earth of suitable radius (k negative) we can, with certain well-known reservations, retain the straight-ray conception. Thus if the slope of BC changes with time, the varying concavity of the effective earth will cause phase changes between ground-reflected and direct waves. On sufficiently short waves this can cause marked fading interspersed with abnormally strong signals approaching or reaching twice the free-space value. The recording of such signal patterns has enabled estimates to be made of the M-gradients within the surface ducts responsible for the effect (Section 2.2.2). The phenomenon is quite distinct from trapping,[17] which occurs extremely rarely in the v.h.f. band.

(2.2) The Estimation of k Variations over Paths of Simple Shape

The principles of Section 2.1 have been applied to an analysis of continuous signal records taken over numerous paths where attenuations in terms of k are calculable with reasonable precision. The path-attenuation calculations, which will not be described in detail, involve two basic principles:

(a) Ray methods well within the visual range, after suitably distorting the path profile to allow for the relevant value of k.

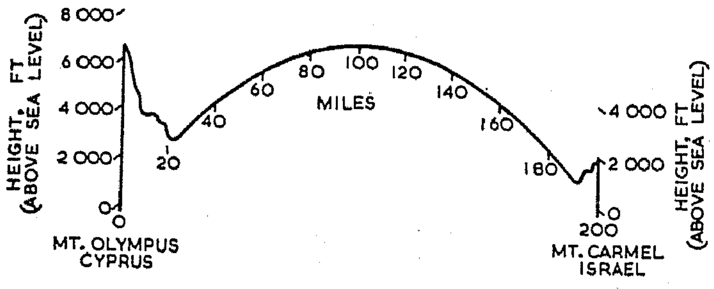


Fig. 5.—Path profile for Israel–Cyprus tests.

(b) The use of the propagation curves of Eckersley and Millington[2] for near-grazing and non-optical paths, after suitable modification of terminal heights and distances to allow for k, and in some cases to allow also for the fact that the path profile had an average radius of curvature differing appreciably from the earth's.

Except for values of k deduced from very extensive signal recordings over the accurately computable Israel–Cyprus path, the k-variations quoted in this Section are not necessarily typical of the regions involved, because they are based on only about 150 hours of record in each case, with no opportunity for choice of weather conditions. All the paths are in the tropics, and it is thought that the figures quoted, even if sometimes untypical, are of scientific and engineering interest. In cases where negative values of k occurred (indicative of a duct), the implicit values of dM/dh and dn/dh have been tabulated for meteorological interest.

In the discussion following, we shall adopt the terms "orthodox signal" and "orthodox path attenuation" to those quantities relevant to propagation in a troposphere of uniform M-gradient, which as previously explained, are obtainable from the signal records after elimination of interference effects from elevated reflecting layers. These orthodox values are then applied to the direct determination of k.

(2.2.1) The Israel–Cyprus Path.

This very long non-optical sea path is worth detailed study in that transmissions on two frequencies were recorded for more than 2 400 hours each, covering the period May–November, 1951, and involving interesting seasonal changes.

C.W. transmitted powers on the two test frequencies of 77·47 and 174 Mc/s were about 200 watts and 3 watts respectively, and Yagi aerials with a gain of 8 dB relative to a half-wave dipole were used at each terminal. Polarization was horizontal.

The path profile is given on Fig. 5. Fig. 6 shows a sample of hourly mean signal-levels on the two test frequencies, and for added interest a scale of k in terms of the 77·47-Mc/s signal is appended. The 174-Mc/s plot may suffer from inaccuracies due to the occasional disappearance of this signal below noise because of the low transmitted power, but correlation with the 77·47-Mc/s plot is good, showing that the charts reflect trends in refraction fairly accurately.

To obtain the variations in k listed in Table 2, orthodox path-attenuations were extracted from the original records by smoothing the variations as already described (Section 2.1) and the values referred to the appropriate curve of theoretical path-attenuation versus k on Fig. 3. Owing to the sometimes prolonged disappearance of the 174-Mc/s signal during the autumn,

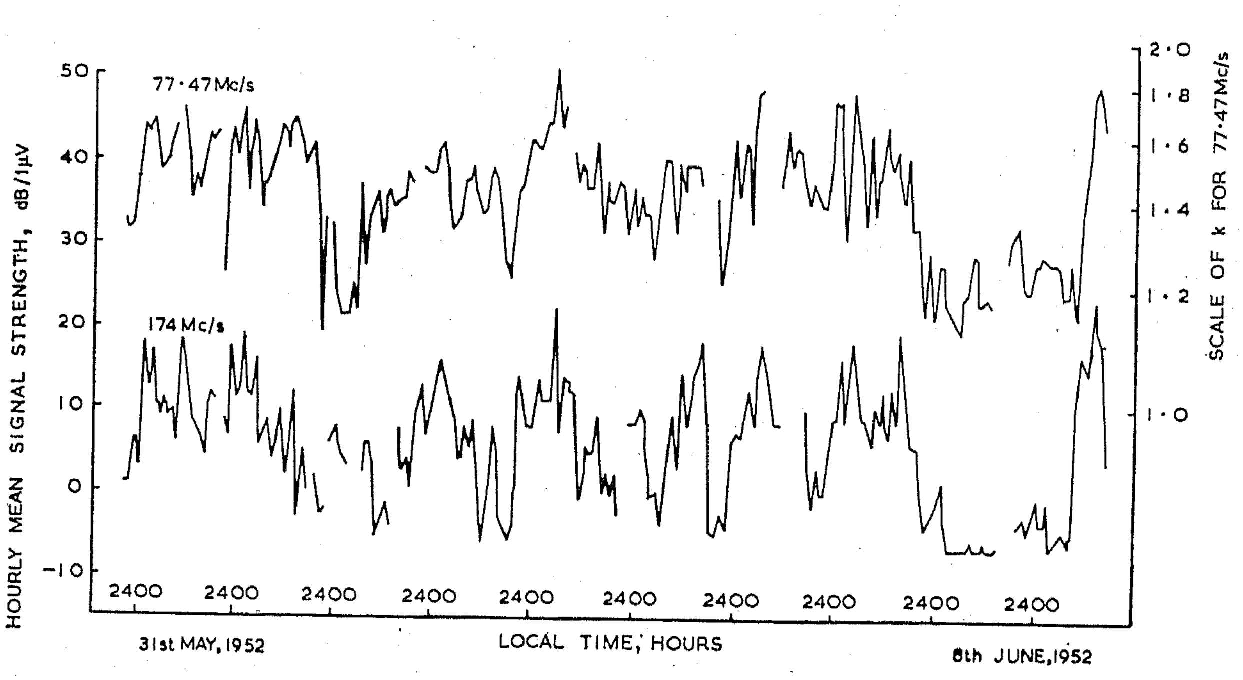


Fig. 6.—Israel–Cyprus path: simultaneous hourly mean signal strengths.

Table 2

	Summer period		Autumn period	
	77·47 Mc/s	174 Mc/s	77·47 Mc/s	174 Mc/s
Minimum orthodox path attenuation	120 dB	118 dB	120 dB	122 dB
Median orthodox path attenuation	139 dB	146 dB	149 dB	153 dB
Maximum orthodox path attenuation	163 dB	Signal below noise	163 dB	Signal below noise
Maximum k	2·0	2·1	2·0	2·0
Median k	1·41	1·44	1·21	1·33
Minimum k	1·0	—	1·0	—

tests on this frequency yielded less reliable values of k than those on 77·47 Mc/s.

The analysis has been divided into a summer period (May–August inclusive) and an autumn period (September–November inclusive). It is seen that on the average there is sufficient reduction of k during the autumn to increase the median path-attenuation on 77·47 Mc/s by about 10 dB. Further statistics of k-variations over this route are given in Section 3.

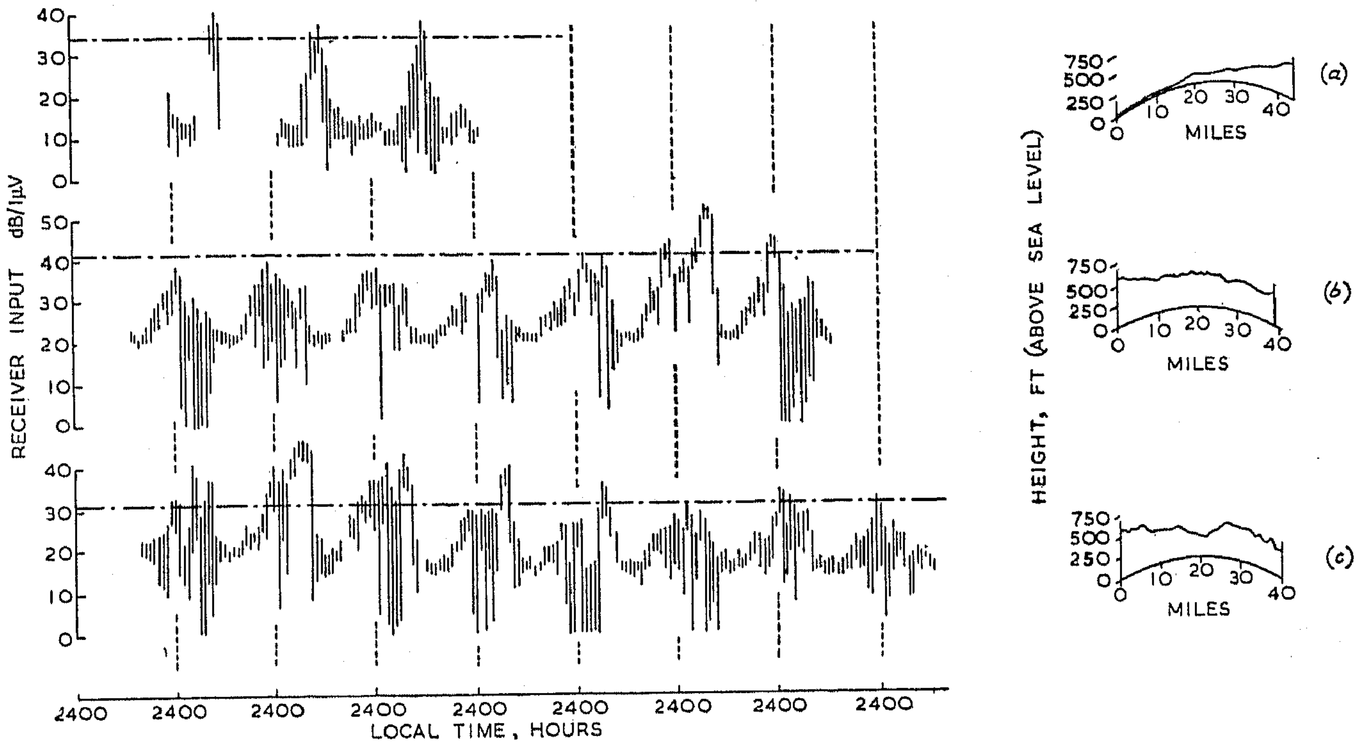


Fig. 7.—West African land paths: hourly signal range charts and profiles.

(*a*) Lagos–Ipara (Nigeria), 174 Mc/s.
(*b*) Abuo–Prang (Gold Coast), 180 Mc/s.
(*c*) Salaga–Prang (Gold Coast), 180 Mc/s.
—·— Free-space signals.

(2.2.2) **West African Land Paths.**

More than 40 land paths have been examined in Nigeria and the Gold Coast, and many of them have exhibited strongly marked diurnal signal variations, usually characterized by steady daytime signals consistent with an approximately "standard" value of k, and very disturbed nocturnal signals attaining abnormally high values, and frequently interspersed with deep fades.

Fig. 7 shows hourly signal ranges and profiles for three land-paths in Nigeria and the Gold Coast exhibiting the effect in a remarkable degree. The extent of the vertical bars shows the maximum and minimum signal strength recorded during any hour. It is noteworthy that the period of orthodox propagation, whose steadiness testifies to a marked absence of tropospheric layering, often starts very abruptly between 0800 and 1000 hours, local time, as is strikingly illustrated by the signal record reproduced in Fig. 8. During early afternoon a steady increase in k usually sets in, culminating in very high (and sometimes even negative) values from about 2000 hours onwards.

During the night we have the situation already discussed (Section 2.1) in which negative or large positive values of k indicate the development of a strongly refracting layer comprising a surface duct or some incipient state thereof. Such surface conditions are, of course, familiar in meteorology, and are promoted by the nocturnal cooling of heated land to a clear sky. The low-level temperature inversion thereby produced is sometimes assisted by a humidity lapse, the combined influences conspiring to produce a strong lapse in refractive index often sufficient to form a duct which is later swept away by morning convection.

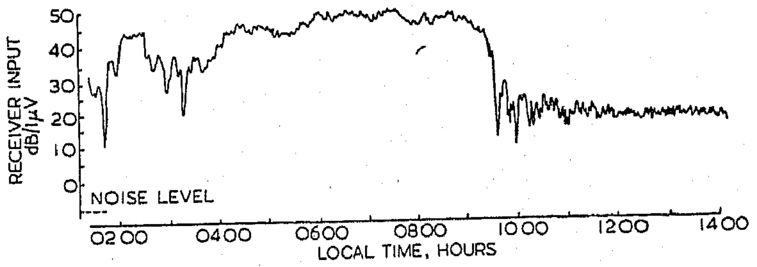


Fig. 8.—Sharp transition from a nocturnal super-refracting to a daytime orthodox condition.

Salaga–Prang (Gold Coast), 180 Mc/s, 12th December, 1952.

The M-curves of Fig. 9 broadly typify the cycle of diurnal tropospheric changes consistent with the signal patterns of Fig. 7. Local meteorological data are unfortunately not available in adequate detail, but the radio behaviour and deduced meteorological structures are analogous with radar performances and surface meteorology over the Bay of Bengal, details of which have been collated by Durst.[10] From October to March, 1943, 200-Mc/s radar sets in Bengal, looking over the damp Ganges delta, experienced inland nocturnal super-refraction for at least

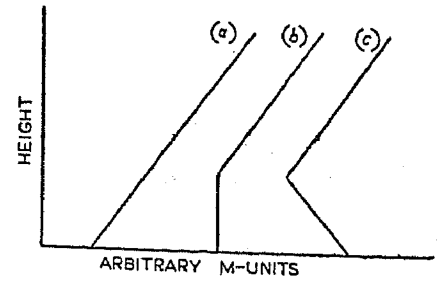


Fig. 9.—Idealized diurnal tropospheric changes implied by West African measurements.

Table 3

Path	Distance	Aerial heights above ground	Frequency	k	$\frac{dM}{dh}$	$\frac{dn}{dh}$	Remarks
	miles	ft	Mc/s				
(*a*)	44	150	174	$-0{\cdot}21$	$-0{\cdot}74$/m	$-0{\cdot}9 \times 10^{-6}$/m	Peak nocturnal
		80		$1{\cdot}3$ to $1{\cdot}4$	—	—	Daytime
(*b*)	39	80	180	$-0{\cdot}16$	$-0{\cdot}96$/m	$-1{\cdot}1 \times 10^{-6}$/m	Peak nocturnal
		80		$1{\cdot}4$	—	—	Daytime
(*c*)	38	80	180	$-0{\cdot}155$	$-1{\cdot}0$/m	$-1{\cdot}2 \times 10^{-6}$/m	Peak nocturnal
		80		$1{\cdot}4$	—	—	Daytime

24 nights a month. Durst[10] quotes the case of a radar equipment whose daytime range of 30–40 miles increased after sunset (March, 1944) to 200 miles. This anomalous range persisted until about 1000 hours next morning.

It is interesting to calculate from the records of Fig. 7 the value of the nocturnal refractive-index gradient, on the assumption that it was constant up to a height exceeding that of the terminals. It is at once noticed that in all three cases a few days' record has yielded a maximum signal significantly exceeding the free-space value marked on the record, which itself is usually about 20dB above the steady daytime level. Within the limits of experimental accuracy, the maximum recorded signals are symptomatic of constructive interference between a direct and a ground-reflected wave of comparable amplitude, a situation only possible in the cases considered when the earth is effectively concave. We therefore know from the outset that, during these periods of exceptional signal, the terminals were lying within a surface duct (k negative). Meteorological measurements show that such ducts are rarely more than 500ft thick, and to ensure, therefore, that the ray paths are confined to a medium of fairly constant M-gradient, we should avoid terminals rising excessively and abruptly above the general ground level if we are to come to valid conclusions about the structure of surface ducts. Although undulating, the three paths considered probably meet this requirement, it being remembered that surface ducts tend to follow small ground irregularities. Further, the results obtained to some extent justify the assumptions of the method.

The observed signal maxima imply that super-refraction occasionally equalled or exceeded the value needed to bring direct and ground-reflected waves into phase. It can be shown that an increase in refractive-index gradient of about a further 40% will bring the waves out of phase and cause a fade. The records indeed show deep fades associated with high nocturnal signal levels, but the theory that the above mechanism is responsible breaks down in most instances when the signal patterns are considered in detail.

Remembering that the intensified super-refraction needed to produce a fade by this mechanism must eventually die away, and in so doing pass through the state of constructive interference, we would expect every fade of this sort to be followed sooner or later by a signal approaching twice the free-space value. Fig. 7 demonstrates, however, that most of the recoveries from numerous nocturnal fades fail to reach the requisite very high level, and we must therefore look elsewhere for an explanation of these fades. They probably stem from some interference effect not considered, due for example to an abrupt change in refractive-index gradient at the top of the duct. We shall therefore ignore these fades from the computational aspect, and calculate on the firmer basis of constructive interference revealed by the occasional presence of signals of about twice the free-space value, with the reservation that the refractive-index gradients so deduced may have been somewhat exceeded.

Table 3 gives for the paths (*a*), (*b*) and (*c*) (see Fig. 7) the values of k (always negative) needed to bring direct and ground-reflected waves into phase. The meteorological quantities listed are of course implicit in k. For added interest, steady daytime values of k are included.

Some comprehensive measurements made over the Arizona desert by Day and Trolese[18] form an interesting parallel with the present observations. Simultaneous v.h.f., u.h.f. and s.h.f. measurements showed varying degrees of nocturnal signal enhancement between the hours of 1800 and 0900. Concurrent low-level meteorological soundings showed:

(*a*) Daytime values of k of about $1{\cdot}2$.
(*b*) Transitions between sub-standard and super-refracting surface conditions at about the times mentioned above.
(*c*) Nocturnal surface ducts extending to 150–200ft and having maximum M-gradients of about $-0{\cdot}15$/m.

Comparison with Table 3 shows the Arizona ducts to be much weaker than those deduced for West Africa. This is confirmed by Day and Trolese's test[18] on 170Mc/s (over a comparable path), which shows a much smaller nocturnal signal increase than the present measurements on virtually the same frequency.

It is of some interest to compare the characteristics of the intense West African ducts with the available world-wide data on tropospheric layers.[19,20,10] It is found that layers can have refractive-index gradients as high as -3×10^{-6}/m, which is much in excess of the West African values (Table 3). The total refractive index change across a layer seldom exceeds 6×10^{-5}, which places a probable limit of about 200ft on the thickness of the ducts investigated. Fig. 10 shows a synthetic M-profile for path (*a*) (Fig. 7) based on the foregoing reasoning.

From the meteorological viewpoint the conditions deduced are quite possible during combinations of a marked temperature inversion with a strong humidity lapse near the ground. Such a conjunction is far more likely in the tropics than in temperate zones because the high nocturnal temperatures in the tropics (usually between 70° and 80°F during the tests) allow a greatly increased humidity lapse rate to develop. Weather observations made during periods of high nocturnal signal usually conformed to this situation; e.g. clear skies, little wind, often followed by morning mist.

It is interesting to compute, from the aspect of possible long-range interference with common-frequency working, the lowest frequency which can be trapped within the deduced ducts listed in Table 3. Taking the values of the duct thickness d and the

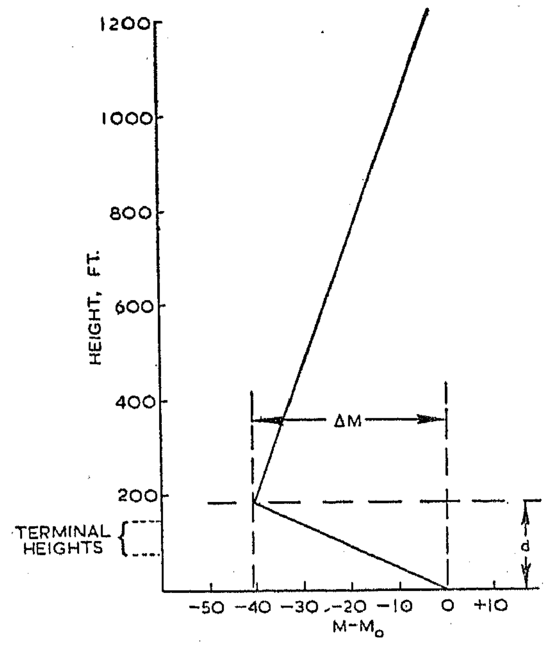


Fig. 10.—Synthetic M-profile consistent with the high nocturnal signals of Fig. 7(a).

M-deficit ΔM given by Fig. 10, it can be shown[17] that trapping in this duct is impossible on frequencies below about 350 Mc/s. It is noteworthy that in Bengal and elsewhere during the war, anomalous ranges were often obtained from radar equipments on frequencies as low as 200 Mc/s.

(2.2.3) West African Coastal Path.

Fig. 11 shows a chart of hourly signal ranges and the profile for an easily computable sea path across a well-indented bay in the Gold Coast. This path, between Cape Coast and Takoradi, appears to be quite free of the diurnal land influences analysed in Section 2.2.2.

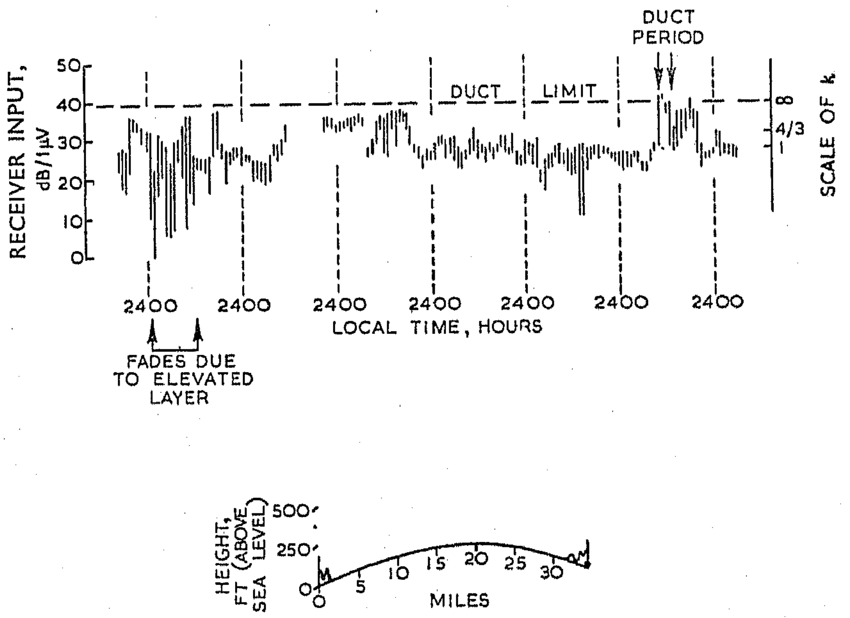


Fig. 11.—Cape Coast–Takoradi (Gold Coast) path: hourly signal ranges on 180 Mc/s, and path profile.

The record indicates a minimum value of k approaching unity. At some periods k-variations are obscured by fades attributable to elevated layer reflections (Sections 2.3 and 2.4). We shall concern ourselves here only with the peak-signal period occurring at noon on one day and identified as a duct in Fig. 11. A scale showing $k = 1$, $k = 4/3$ and $k = \infty$ (the duct limit) has been added to the chart, and Table 4 gives the maximum meteorological values computed for the duct.

The figures indicate a rather weak duct, which it is interesting to compare with a representative coastal sounding made in this region in 1927 on the S.S. *Meteor*[12] during a meteorological survey in the Atlantic. Fig. 12, representing *Meteor* sounding 182, shows a more intense duct than that implied by the present radio measurements, but it illustrates a kindred condition.

Table 4

Distance	Aerial heights above sea level	Frequency	k	$\frac{dM}{dh}$	$\frac{dn}{dh}$
miles	ft	Mc/s			
34·5	250 260	180	−5·0	−0·031/m	$-0{\cdot}19\times10^{-6}$/m

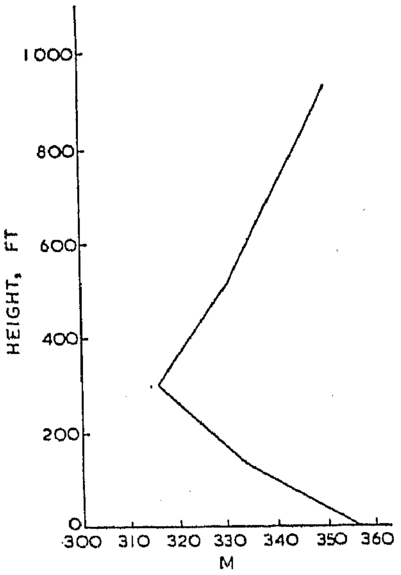


Fig. 12.—M-profile from S.S. *Meteor* sounding No. 182.

(2.2.4) Paths over Lake Victoria.

In 1951 and 1952 tests were made on a 51-mile optical path over Lake Victoria, East Africa (Fig. 13), which exhibited destructive interference between direct and water-reflected waves whenever k passed through a critical value.

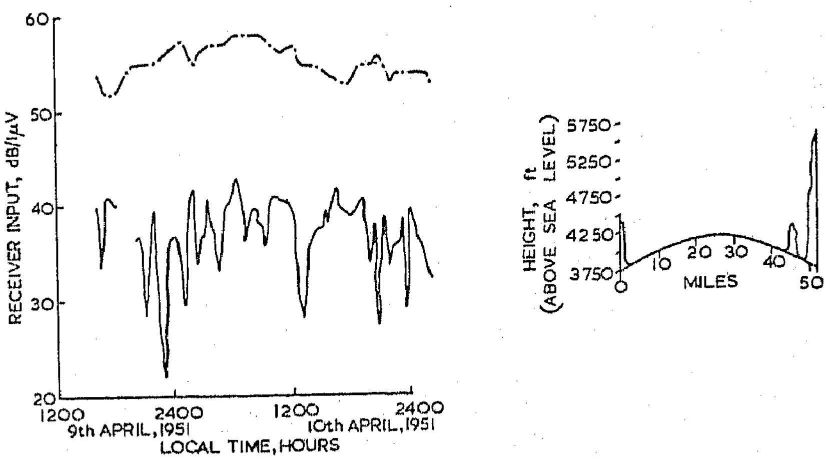


Fig. 13.—Gemba–Sigulu (Lake Victoria) path: simultaneous signal
—·—·— 77 Mc/s.
———— 169 Mc/s.

When variations in k cause a modification in the effective earth radius, the effective terminal heights for an optical path are altered and may thus bring waves travelling the direct and ground-reflected paths into phase opposition, with consequent fading. This only happens when the effective path difference is any multiple of a wavelength. For most optical paths subject to normal variations in k (1 to 2, say), the path difference is less than λ in the v.h.f. range, and fading due to this cause is thus rare. For the path in question, however, the terminals were sited very high above the good reflecting surface of the lake, and this type of fading often occurred on one of the test frequencies.

In May, 1951, simultaneous tests were made on 77 Mc/s and 169 Mc/s, using powers of about 8 watts and horizontal Yagi aerials. Fig. 14 shows the theoretical relationship between k and the received signal computed for the two transmissions. The calculation assumes a reflection coefficient of -1 for horizontally polarized waves at grazing incidence to water, and neglects the effect of ray divergence at a curved surface. The higher frequency is seen to suffer cancellation whenever k passes through 1·46, while the lower frequency is insensitive to variations in k.

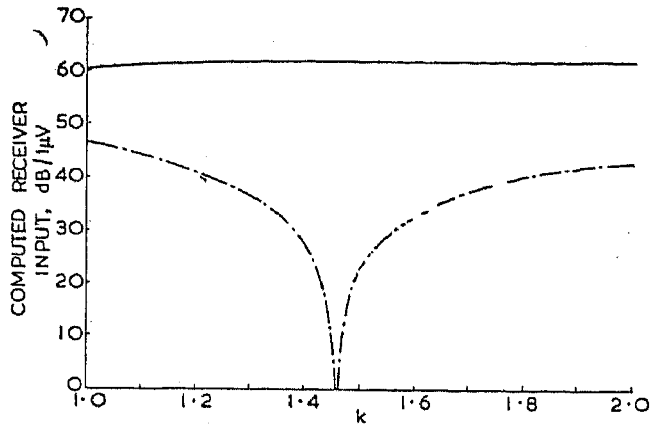


Fig. 14.—Gemba–Sigulu path: theoretical variation of signal strength with k for the two test frequencies.

——— 77 Mc/s.
—·—·—· 169 Mc/s.

These expectations are borne out by Fig. 13, which shows a short signal record of the two transmissions. As expected, the higher frequency suffers deep fading while the lower frequency is relatively immune. The test was repeated in April of the following year on the higher frequency only, with similar results. We conclude, therefore, that k frequently reached or exceeded 1·46, but the actual excess cannot be determined. As the weather was often turbulent during the tests, it is probable that high values were not reached. Table 5 includes details of the test.

Table 5

Path length over lake	Aerial heights above lake	Frequency	Minimum k	Median k	Maximum k
miles	ft	Mc/s			
51 (optical)	700 1 800	169	1·1	—	>1·46
92 (non-optical)	740 1 800	77 169	0·9 1·0	1·2 1·2	1·9 1·6

The fading exemplified here can be eradicated in the v.h.f. range (for normal variations in k) by lowering the terminal heights or reducing the operating frequency. When these measures are impossible, the depth of fading can theoretically be reduced by the use of vertical polarization, because the reflection coefficient for water, even near grazing incidence, is then significantly less than unity owing to the very small pseudo-Brewster angle in the v.h.f. range. This fact was overlooked at the time of the tests, in which horizontal polarization was used. For horizontal polarization the depth of fading should be governed mainly by a divergence factor appropriate to the effective earth-radius producing the fades. Possibly because of partial reflection from elevated layers, the observed depth of fading often considerably exceeded the theoretical figure.

Further tests in April, 1951, were made on an adjacent 92-mile non-optical path, again substantially confined to Lake Victoria. In this case the expected signals on two test frequencies were computed in terms of k, using the propagation curves of Eckersley and Millington[2] as mentioned in Section 2.2. Variations in k were then deduced in the way already described, from a 9-day simultaneous recording on 77 Mc/s and 169 Mc/s. Table 5 includes details of the test.

The discrepancy in the maximum values of k for the longer path is due in part to the steepness of the curve of signal versus k in this region. The relatively small range of k is consistent with the figures obtained from the adjacent optical path previously discussed, and is attributable to the atmospheric turbulence often occurring in this upland region.

(2.3) The Measurement of Tropospheric Reflection Coefficients from Recorded Fades

Over most v.h.f. paths, fading is attributable to partial cancellation of the orthodox signal by a signal partially reflected from an elevated atmospheric discontinuity, typified by Fig. 1(*a*). This fact has enabled estimates of tropospheric reflection coefficients to be made from analysis of recorded fades in simple cases.

For a fade of appreciable depth, orthodox and tropospherically reflected waves are, of course, nearly equal in amplitude (presuming phase opposition), and hence knowledge of orthodox signal amplitude determines that of the tropospheric signal. When this signal is mainly attributable to a single dominant ray-path via an elevated layer, we can estimate the modulus of the reflection coefficient of the layer.

Calling the layer reflection coefficient ρ, its modulus $|\rho|$, and E_{pr} the free-space signal, we have a tropospheric signal ρE_{pr}, which during a fade partly annuls the orthodox signal which we shall call E_0.

On the assumption of momentary phase opposition at the bottom of the fade, the signal will fade to a value

$$|E_0| - |\rho|\,|E_{pr}|$$

namely, to $\left[1 - |\rho|\dfrac{|E_{pr}|}{|E_0|}\right]$ of its orthodox value.

Calling this fraction $1/N$, we derive

$$|\rho| = \frac{|E_0|}{|E_{pr}|}\left[1 - \frac{1}{N}\right]$$

For fades exceeding about 20 dB, for which $N > 10$, $|\rho|$ is determined to an adequate approximation by $|E_0|/|E_{pr}|$.

In simple cases $|E_0|$ can be approximately estimated from a record of a simple fade, making the following assumptions:

(*a*) That shortly before or after the fade reflected and orthodox waves come into phase.

(*b*) That $|E_0|$ and $|\rho|$ remain fairly constant during the complete fading cycle.

(*c*) That the reflected wave results mainly from a single tropospheric ray path.

It has already been mentioned (Section 2.1) that tropospherically reflected waves can undergo large phase changes with only small modifications to the layer structure and consequent value of $|\rho|$. Further, the orthodox signal $|E_0|$ is likely to remain fairly constant over the fading period, being governed by slow refraction changes. Thus conditions (*a*) and (*b*) are often satisfied, and the likelihood of their fulfilment is best judged by examination of fading records.

Fig. 15 shows a section of signal record suitable for computation, taken over the Israel–Cyprus path. Assuming aiding phase at A and B, the orthodox signal E_0 will lie as indicated, approximately 6 dB below the level of A and B. The ratio N is given by

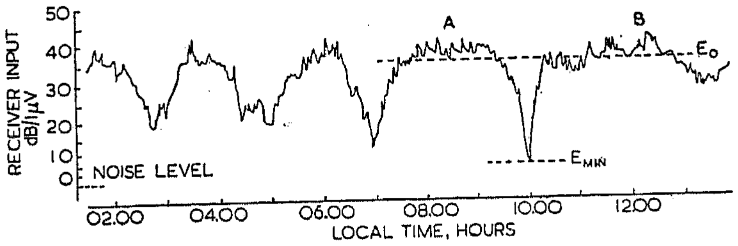


Fig. 15.—Fades suitable for estimation of tropospheric reflection coefficients.

Israel–Cyprus, 77·47 Mc/s, 17th July, 1952.

E_0/E_{min}, which in this case is about 30. To a good approximation, $|\rho|$ is $|E_0|/|E_{tr}|$ in this instance.

The possibility of multiple tropospheric-ray paths involving ground reflection must be considered. These theoretically vitiate assumption (*c*), but fortunately the critical dependence of $|\rho|$ on the incident angle to the layer (Section 2.1) makes it often happen that only one ray path (that involving the minimum grazing angle) is of any consequence in producing a tropospheric signal. Thus assumption (*c*) is often justified to an adequate approximation.

Table 6

Path	Frequency	Date measured	Orthodox path attenuation	Fading depth below orthodox signal	Deduced reflection coefficient $\|\rho\|$
	Mc/s		dB	dB	
Israel–Cyprus	77·47	6th June, 1952	130	7	0·11
	77·47	7th June, 1952	130	13	0·16
	77·47	14th July, 1952	138	36	0·09
	77·47	8th August, 1952	129	17	0·20
	77·47	9th August, 1952	129	18	0·20
	77·47	13th August, 1952	129	20	0·20
	174	25th October, 1952	129	24	0·47
	174	26th October, 1952	129	14	0·40
	174	30th October, 1952	136	20	0·20
Cape Coast–Takoradi . .	180	5th February, 1953	122	28	0·20
	180	5th February, 1953	119	23	0·26

Israel–Cyprus: Path length, 200 miles. Terminal heights, 1 800 ft and 6 470 ft above sea level. Free-space path attenuations, 116 dB on 77·47 Mc/s and 123 dB on 174 Mc/s.
Cape Coast–Takoradi: Path length, 34·5 miles. Terminal heights, 260 ft and 250 ft above sea level. Free-space path attenuation, 108 dB.

The reflection coefficients listed in Table 6 have been calculated from well-defined and simple fades selected from recordings over two water paths already discussed from other aspects, namely Israel–Cyprus and Cape Coast–Takoradi (Gold Coast). For convenience the relevant variables are quoted in terms of attenuation (decibels). The values of $|\rho|$ must be considered as exceptional rather than typical, as the aim has been (*a*) to select outstanding fades indicative of well-defined stratified layers, and (*b*) to calculate the meteorological implications of some of the largest reflection coefficients observed.

Table 6 shows that we must occasionally expect very large reflection coefficients from elevated layers, and, as Fig. 4 demonstrates, these coefficients require very grazing angles of ray incidence for possible layer structures. Other water paths in various tropical regions have given figures comparable with those listed.

Further work is necessary to obtain reliable figures for land paths. Of the available records of land paths showing fades, which seem initially to testify to layers aloft, many fade only at night from abnormal levels indicative of a simple surface duct rather than an elevated layer (Section 2.2.2). These fades, which are not convincingly explainable in terms of a simple ray picture, are not valid material for the present purpose. The small fading exhibited by suitably qualified land paths [see Fig. 24(*a*)] implies small values of $|\rho|$ and hence weak stratification. The present limited data suggest that elevated stratification is appreciably less pronounced over land than over water.

(2.4) Meteorological Implications of Observed Tropospheric Reflection Coefficients

The first complete day's test of the Cape Coast–Takoradi path (Fig. 11) showed fades indicating the presence of an elevated layer, the derived reflection coefficients of which have been listed in Table 6. It is of some interest to deduce, with the assistance of charts of tropospheric reflection coefficients due to Millington,[16] the layer heights compatible with these coefficients.

We shall take a representative layer thickness of 200 ft, and the probable limits of Δn, the change of refractive index across the layer, as 10^{-5} and 5×10^{-5}. The large values of $|\rho|$ under investigation (0·20 and 0·26) show that the reflection angle at the layer (presumed above the terminals) only slightly exceeded the critical angle for total internal reflection. As this angle depends only on Δn, uncertainties in the layer thickness will not greatly affect the results. We shall make use of the reflection-coefficient curves of Fig. 4 previously discussed, which are appropriate to the chosen layer thickness and the test frequency of 180 Mc/s.

Assuming that the equal terminal heights produced a reflection at mid-path from a spherically stratified layer, we can plot, with the aid of Fig. 4, a curve of Δn versus the layer height compatible with the observed value of $|\rho|$. In this process the layer curvature must be adjusted to conform to the effective earth-radius factor k, operative at the time of the fade. This value of k is calculable from the record, as explained in Section 2.1.

Fig. 16 shows curves derived by applying this process to the two values of $|\rho|$ already noted, making due allowance for the relevant values of k. The two observations, which were made eight hours apart, lead to very similar conclusions.

Numerous soundings made on the S.S. *Meteor*[12] mentioned in Section 2.2.3, indicate that, in the region of the test, Δn is likely to have an average value of about 3×10^{-5} and a maximum value of about 5×10^{-5}. A few of the *Meteor* observations showed layers barely qualifying as a duct. For the layer thickness of 200 ft this places the probable lower limit of Δn at about 10^{-5}. Reference to Fig. 16 gives the corresponding range of layer heights as 500–1 100 ft, with 850 ft as the most probable value.

This deduced range of possible layer heights, coupled with the implied appearance of a layer on the surface at another time during the same radio test (Section 2.2.3), is in general conformity with coastal meteorology, where a transition from a surface duct

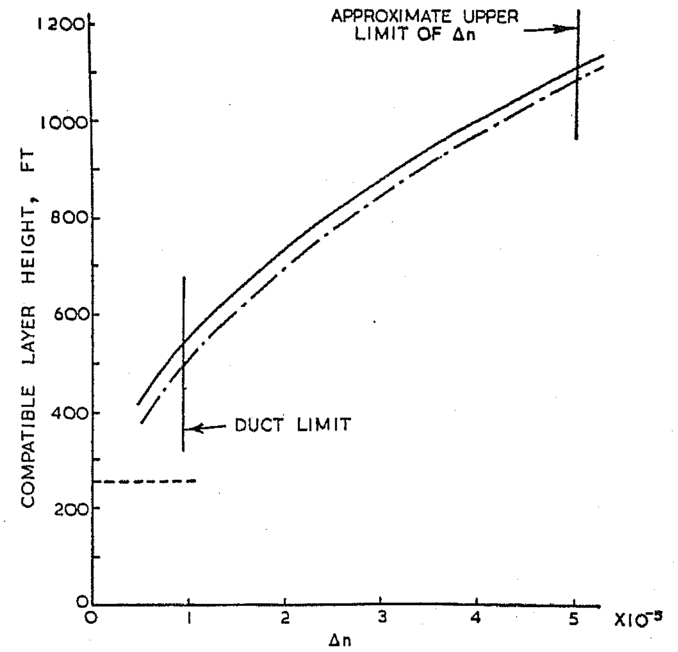


Fig. 16.—Cape Coast–Takoradi path: relationships between Δn and layer height compatible with radio measurements on 180 Mc/s.

——— $|\rho| = 0{\cdot}26$, $k = 1{\cdot}4$.
—·—·— $|\rho| = 0{\cdot}20$, $k = 1{\cdot}0$.
— — — Height of terminals.

(Fig. 12) to a low elevated duct such as deduced above, is quite common.

(3) SOME STATISTICAL ASPECTS OF V.H.F. PROPAGATION

(3.1) The Form of Presentation of the Results

This Section summarizes from an engineering viewpoint the statistical behaviour of the numerous v.h.f. radio paths mentioned in Section 1. The material is derived (*a*) from about 15 000 hours of signal records taken over tropical paths, individually recorded for about 150 hours each, and (*b*) from intermittent recordings of transmissions between Israel and Cyprus totalling over 5 000 hours, over the period May–November, 1952. The frequencies employed for the tests were usually near 170 Mc/s, but in a few cases frequencies near 80 Mc/s were used, sometimes simultaneously with a higher frequency. Path lengths varied between about 12 and 200 miles, averaging about 40 miles, while terminal heights ranged between 50 ft and 10 000 ft above sea level. Only 40% of the paths were optical, while 15% were predominantly over water. It is the purpose of this Section to compress the analysis of these measurements into a form useful to the communication engineer. The results are mainly applicable to tropical regions, but it is thought that they might provide a framework for the approximate estimation of v.h.f. behaviour in temperate climates.

(3.2) Types of Signal Variations

Signal fluctuations due to changing atmospheric conditions will generally increase with the separation of the terminals. In the v.h.f. range, variations are usually negligible at distances below 25 miles. The signal patterns discussed below have been found characteristic of distances ranging between 35 and 200 miles. The characteristics of diurnal, seasonal and random variations will be summarized separately, and the number of paths observed in each category will be quoted to give a rough idea of the significance of the conclusions.

(3.2.1) Diurnal Variations.

These variations are divisible into three categories exemplified in Fig. 17:

(*a*) Small diurnal fluctuations devoid of fades.
(*b*) Large diurnal variations characterized by very steady signals round mid-day, and violently fluctuating signals nocturnally.
(*c*) Diurnal variations superficially resembling (*b*), but distinguished by irregular daytime signals.

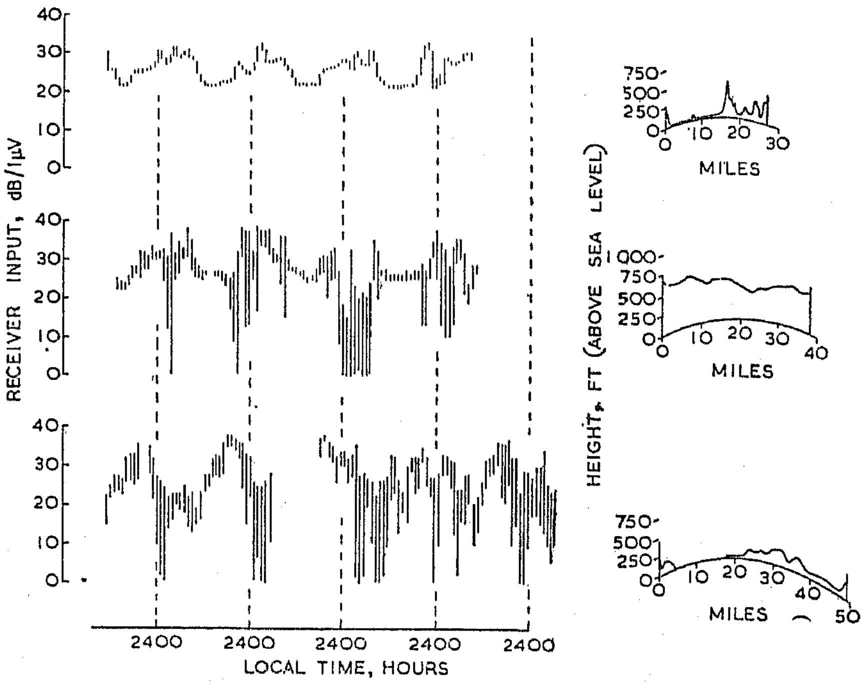


Fig. 17.—Representative types of diurnal signal variations.

(*a*) Malaya (Telok Datoh–Kuala Lumpur, 175 Mc/s).
(*b*) Gold Coast (Tamale–Palbusi, 180 Mc/s).
(*c*) Gold Coast (Cape Coast–Winneba, 180 Mc/s).

For ease of comparison, examples of these three types are plotted as hourly signal ranges in Fig. 17.

Type (*a*) is representative of 26 paths measured in Malaya and eight paths in Ceylon. The absence of fades indicates an unstratified atmosphere, and the smooth diurnal signal variations reflect small diurnal variations in atmospheric refractive-index gradient.

Type (*b*), of which 20 cases were found inland in West Africa, is well exemplified in Fig. 7 previously discussed, but another example is shown in Fig. 17 for comparison purposes. This signal pattern is characteristic of large land masses subject to nocturnal cooling, when at least one of the terminals is low enough to lie within the resulting nocturnal surface duct (Section 2.2.2).

Type (*c*), of which four examples were found along the West African coast, appears to be characteristic of grazing coastal paths predominantly over sea yet never departing more than a mile or two from a large land mass. As might be expected, there are points of resemblance with inland paths, but the coastal paths show significantly greater signal variations round midday. These irregularities are attributable to interference effects resulting from changes in the height of the base of low coastal ducts, which, unlike inland surface ducts, can exist by day and are governed by alternating land and sea breezes. Furthermore, types (*b*) and (*c*) usually show a significant difference in the time of occurrence of the nocturnal maximum signal.

(3.2.2) Seasonal Variations.

The Israel–Cyprus path (see Fig. 5) was the only route tested long enough to reveal seasonal changes. Fig. 18 shows the

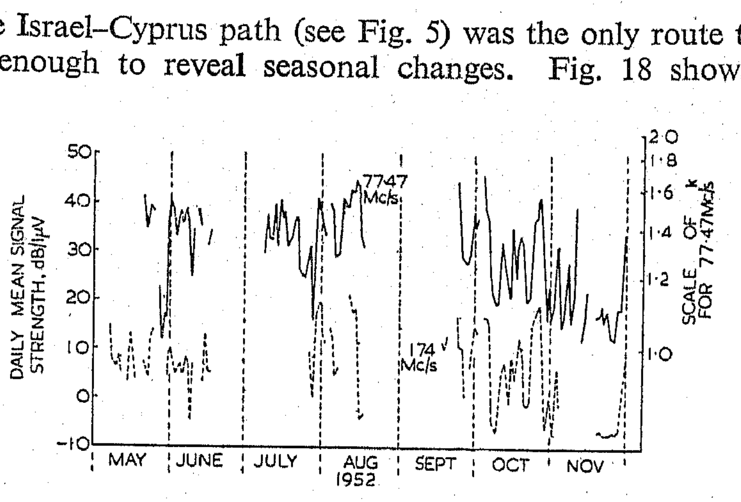


Fig. 18.—Israel–Cyprus path: simultaneous daily mean signal strengths.

variation throughout the test period of daily mean signals on two differing frequencies. Although high signal-strengths occurred frequently throughout the test, there was a noticeable downward trend in the level as the autumn progressed. Fig. 19(*a*) shows the percentage of the time that a specified path-attenuation on 77·47 Mc/s was exceeded during the summer (May to mid-August), and during the autumn (mid-September to November). It is noteworthy that while autumn shows a 12-dB increase in median path-attenuation, the chances of a very high or very low path-attenuation at this season about equalled the chances of such path-attenuations in summer. Simultaneous measurements on 174 Mc/s led to similar conclusions. Fig. 19(*b*) shows similar cumulative distributions of path attenuation for the whole 7-month test period, both for 77·47 Mc/s and 174 Mc/s. It is unlikely that these distributions differ greatly from what would have been obtained from a whole year's measurement; and, pending further data, Fig. 19(*b*) may be taken as broadly representative of tropospheric conditions over the Mediterranean Sea. It should, however, be noted that, owing to occasional loss of the signal below noise, the 174-Mc/s results are less reliable than those relating to 77·47 Mc/s.

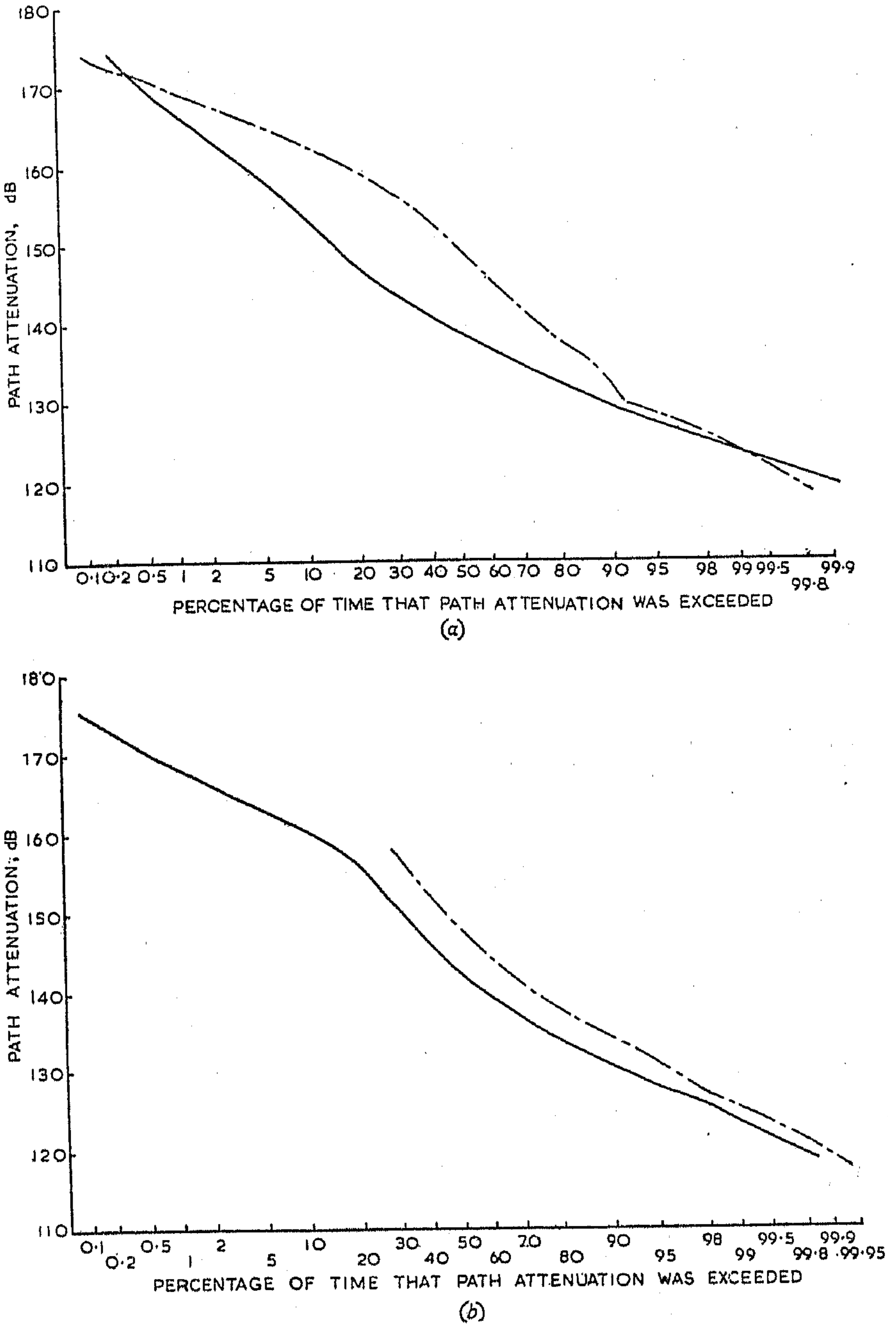


Fig. 19.—Israel–Cyprus path: cumulative distributions of path attenuation.

(*a*) 77·47 Mc/s during summer and autumn.
——— Summer.
—·— Autumn.

(*b*) 77·47 Mc/s and 174 Mc/s during the whole test period.
——— 77·47 Mc/s.
—·— 174 Mc/s.

(3.2.3) **Random Variations.**

Figs. 6 and 11, discussed from other aspects in Section 2, illustrate signal variations typical of water paths where the line of sight is well clear of coast-lines for most of its length. This type of pattern was yielded by two long Mediterranean paths, a non-optical path over Lake Victoria (East Africa), and an optical path across a well-indented bay in the Gold Coast. Random fading characteristic of all these paths shows them subject to pronounced stratification, but they appear quite free of the diurnal influences associated with inland and coastal paths.

Although classified as random, these signal patterns exhibit trends governed by changing atmospheric refraction related to the weather and the seasons. Fig. 20 shows three samples of signal records from the Israel–Cyprus path, illustrating typical behaviour of two simultaneously transmitted frequencies under conditions of:

(*a*) Anticyclonic weather, associated with temperature-inversion layers above a region of fairly high refractive-index gradient.

(*b*) Transition from anticyclonic to turbulent weather.

(*c*) Turbulent weather, associated with a sub-standard refractive-index gradient, where at large distances the weak orthodox signal is vulnerable to scattering from small tropospheric irregularities, with consequent scintillation.

In the random category we should also include the fading sometimes resulting from phase opposition between direct and ground-reflected waves at critical values of k (Section 2.2.4). An example of this type of fading occurs in Fig. 13, and a second example (not illustrated) was found over a land path of appropriate form in Malaya.

3.3) Statistics of k Variations and their Application to Prediction of Path Behaviour

In so far as we can relate measured path attenuations to k (Section 2.2), we can derive a chart showing the percentage of the time that a given value of k was exceeded during the period of test of a specific path. Such a chart is of assistance in predicting the behaviour of other radio paths subject to similar atmospheric conditions, provided that the paths in question are amenable to reliable signal-strength computation in terms of k. To demonstrate this principle the statistics of Fig. 19(*b*) are used below to predict the probable statistical behaviour of a fictitious, but representative, v.h.f. radio path presumed subject to the appropriate atmospheric conditions. Such predictions should be applicable, for example, to paths over the Mediterranean Sea, where the test yielding Fig. 19(*b*) (Israel–Cyprus path) was made.

A distribution of path attenuation such as Fig. 19(*b*) is influenced for a small percentage of the time mainly by interference fades resulting from reflections from elevated tropospheric layers, while for larger percentages of the time the distribution is dominated by variations in k. The boundary between the two domains is somewhat indefinite, but we shall initially assume that the influence of k is dominant so long as k is greater than unity. On this basis we can, with the assistance of the theoretical curves of Fig. 3, relate any value of k between 1 and 2 to a specified attenuation over the Israel–Cyprus path. The application of these path attenuations to Fig. 19(*b*), which is derived from tests over the above route, yields Fig. 21(*a*), which shows for the whole test period the percentage of the time that a given value of k was exceeded over this route. For interest, curves based on each of the two test frequencies are given. Ideally they should be coincident. The discrepancy is surprisingly small considering the errors likely to result from the idealizations implicit in the method. Because of the occasional loss of the 174-Mc/s

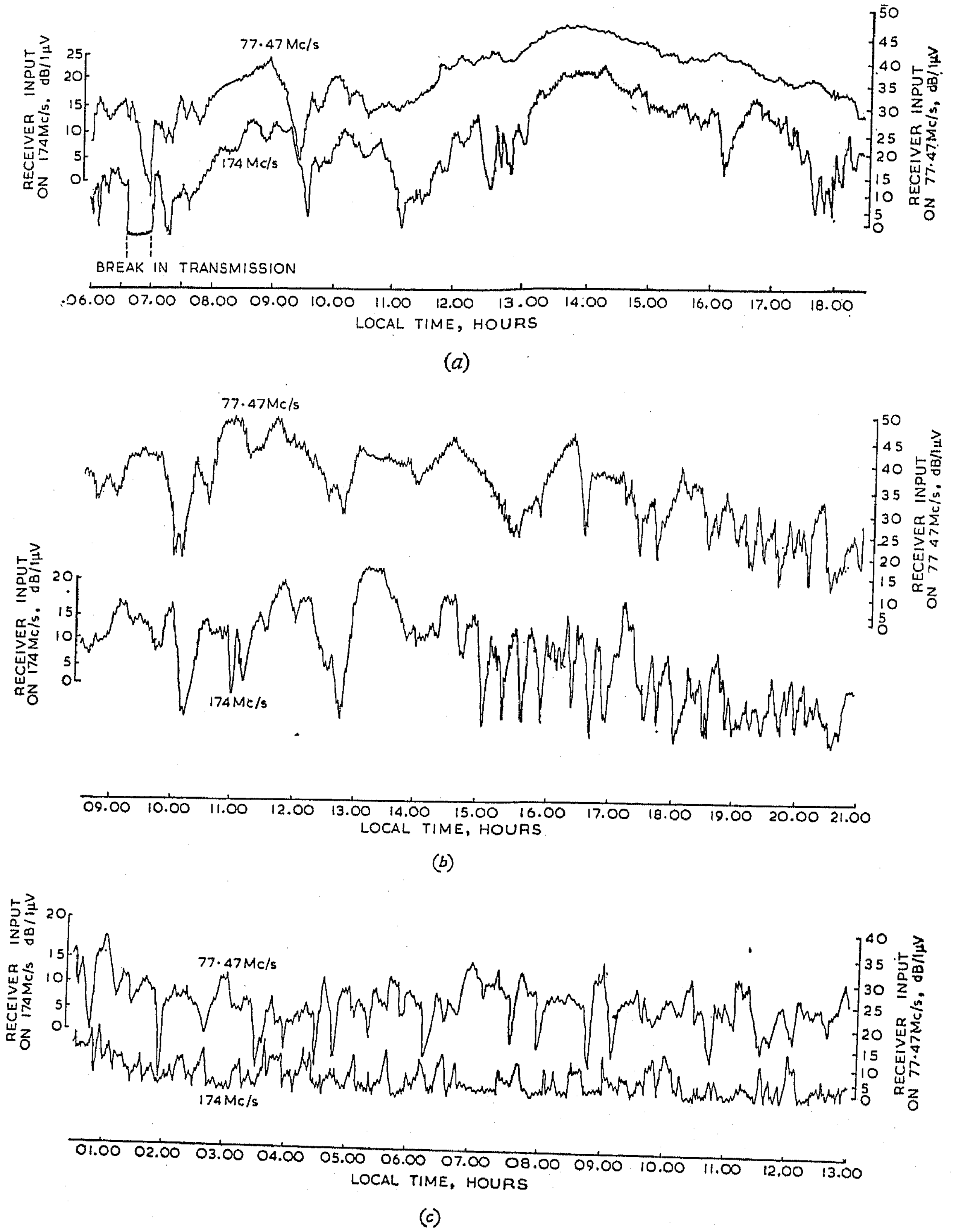


Fig. 20.—Characteristic signal patterns.

(a) Anticyclonic weather; 30th May, 1952.
(b) Transition from anticyclonic to turbulent weather; 7th June, 1952.
(c) Turbulent weather; 8th June, 1952.

(Selected from Israel–Cyprus records.)

signal below noise, the 77·47-Mc/s curve (full line) should be considered as the more authentic. On this basis the median value of k for the whole test period was 1·34, which is virtually the recognized "standard" value.

Fig. 21(*b*) shows distributions of k similarly computed for summer and autumn separately, and based solely on the 77·47-Mc/s measurements over the same path. The median values of k for summer and autumn were respectively 1·41 and 1·20.

Owing to the enhanced relative importance of elevated layer reflections as k decreases, the curves of Fig. 21 may be unreliable very near $k = 1$, but it is shown in Section 3.3.1 how, from the aspect of predicting path behaviour, this limitation can be made irrelevant by the supplementary use of fading statistics.

(3.3.1) The Prediction of Path Behaviour from the Statistics of k Variations.

To demonstrate how the curves of Fig. 21 can be used to predict the behaviour of any radio path of simple profile subject to appropriate tropospheric conditions, we shall forecast the long-term statistical behaviour on 150 Mc/s of a fictitious 50-mile over-water path having terminals 500 ft above sea level. We shall assume the path subject to the tropospheric conditions operative during the long-term test from which Fig. 21(*a*) was derived. It is considered that such conditions are approached over much of the Mediterranean Sea.

Fig. 22(*a*) shows the theoretical variation of path attenuation on 150 Mc/s as k changes from 1 to 2. Combining this curve with the long-term distribution of k given in Fig. 21(*a*), we derive

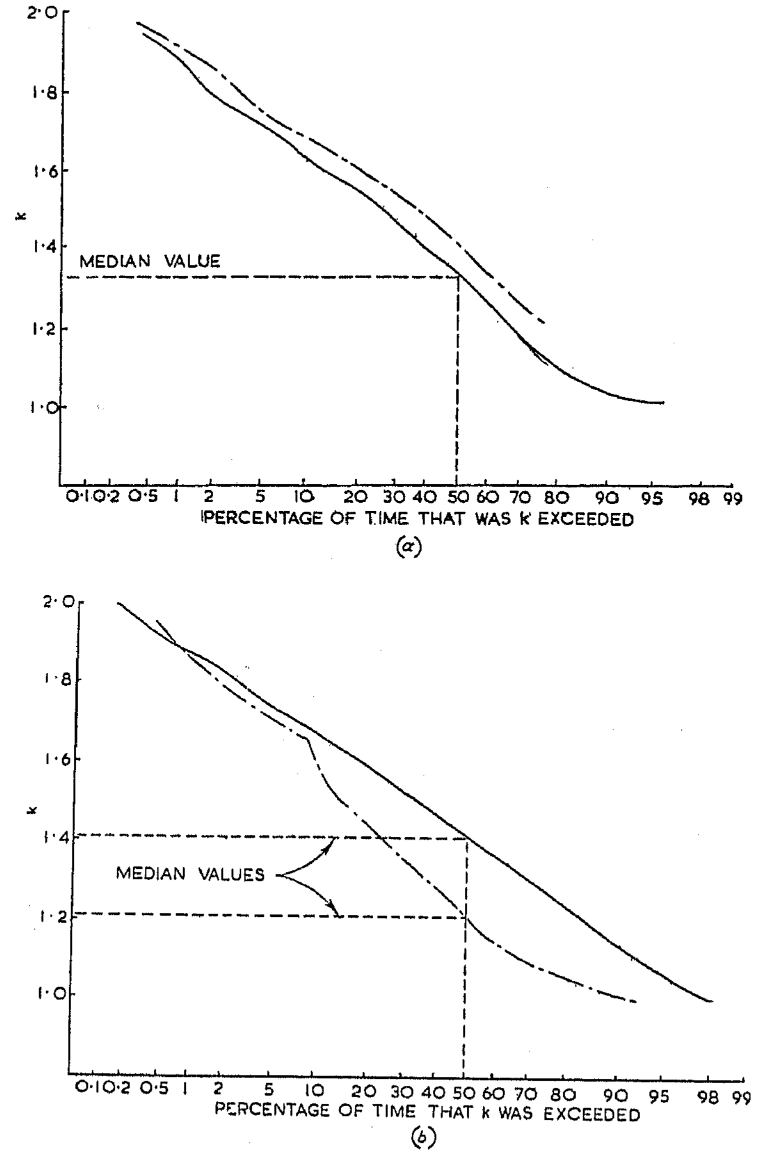


Fig. 21.—Israel–Cyprus path: cumulative distributions of *k*.

(*a*) Whole test period, based on both test frequencies.
——— 77·47 Mc/s.
—·— 174 Mc/s.

(*b*) Summer and autumn, based on the 77·47-Mc/s test.
——— Summer.
—·— Autumn.

the full-line curve of Fig. 22(*b*), which shows the expected path-attenuation distribution derived solely from the expected variations of *k*. As implied in Section 3.3, layer reflections, rather than *k*-variations, exert a dominant influence on the large path-attenuations which are exceeded for small percentages of the time. Thus the latter curve, by virtue of its derivation purely from variations in *k*, must be in error at its low-percentage end.

To correct and extend the left-hand end of the curve we require fading statistics in the form of path attenuations exceeded for some convenient small percentage of the time. From the engineering aspect 0·1% of the time (about $1\frac{1}{2}$ minutes a day) is a useful value. It is shown in Section 3.4 that for v.h.f. paths subject to comparable climatic conditions, there is, with various reservations, a relationship between the path attenuation exceeded for 0·1% of the time and the median path-attenuation. Hence an estimate of the median path-attenuation of a prospective v.h.f. route enables an approximate forecast to be made of the path attenuation which is likely to be exceeded for 0·1% of the time, so long as the route is subject to tropospheric conditions similar to those governing the tests on which the fading statistics are based.

From Fig. 22(*b*) the expected median attenuation of the fictitious path is seen to be 117 dB, and the fading range chart for water paths [Fig. 24(*c*)] shows that this median attenuation will most probably be associated with a fading range of about 12 dB, which, by definition (Section 3.4), will result in a path attenuation of about 129 dB being exceeded for 0·1% of the time. This provides a supplementary point A on Fig. 22(*b*), which enables us to extend and correct the inaccurate end of the full-line curve based on variations of *k*.

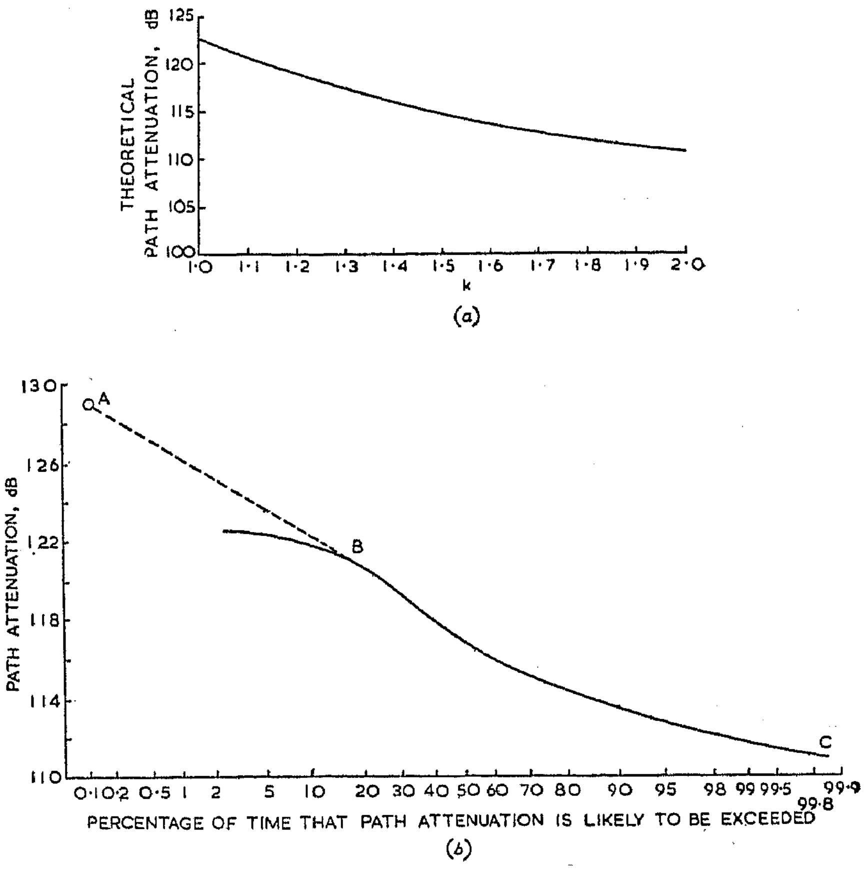


Fig. 22.—Fictitious oversea path.

(*a*) Theoretical variation of path attenuation with *k* for 150 Mc/s.
(*b*) Forecast of cumulative distribution of path attenuation for 150 Mc/s.

To interpolate between the point A and the curve, which is plotted on Gaussian probability paper, a straight line is arbitrarily drawn passing through A and tangential to the curve at B. The forecast distribution of path attenuation is then given by the composite line ABC.

The forecast presumes a Gaussian distribution for small percentages of the time, and, although lacking theoretical justification, this assumption has been found approximately true for long paths. For shorter paths, where elevated layer reflections are relatively too weak to cause deep fading, the curve tends to flatten (on a Gaussian scale) with decreasing percentages of the time. The above simple engineering solution is, however, offered as likely to provide a useful approximate forecast of the actual path performance.

(3.4) A Relationship between Median Path-Attenuation and Fading Range

It has been suggested that under specific meteorological conditions and within a restricted frequency band, there might be a unique relationship between the depth of fading experienced over any v.h.f. radio path and its median attenuation. Such a relationship, if found to apply to a useful approximation, would be of value in forecasting the statistical behaviour of untested v.h.f. paths or paths tested for an inadequate time.

From simple interference considerations, the depth of fading over any v.h.f. path depends solely on the relative amplitudes of a ground wave and a wave of fortuitously opposing phase reflected or scattered from one or more elevated tropospheric layers. By ground wave is meant the wave controlled by the tropospheric stratum between the ground and the lowest layer, which often has an approximately constant refractive-index gradient over this height interval. From this starting-point it can be shown that the suggested relationship should hold if the following conditions are met:

(*a*) The ground-wave amplitude governs the median path-attenuation.

(*b*) The amplitude of the interfering tropospheric wave is uninfluenced by the path profile.

(*c*) The tropospheric wave amplitude is independent of the path length.

(*d*) The tropospheric wave amplitude is independent of frequency.

It is worth briefly examining how nearly these conditions are likely to be fulfilled in practice.

For paths having a commercially tolerable degree of fading, condition (*a*) will hold to a fair accuracy. Condition (*b*) is likely to hold for most practical paths where screening of the terminals by local obstructions must in any case be avoided. Condition (*c*) is the least likely to be met, but it can be shown to hold approximately at sufficient range for a horizontally stratified layer of given height, for the reason that, beyond a certain range, the reflected field from such a layer can remain fairly constant with distance by virtue of the counterbalancing influences of increasing distance and increasing reflection coefficient (the latter resulting from decreasing grazing angle at the layer).

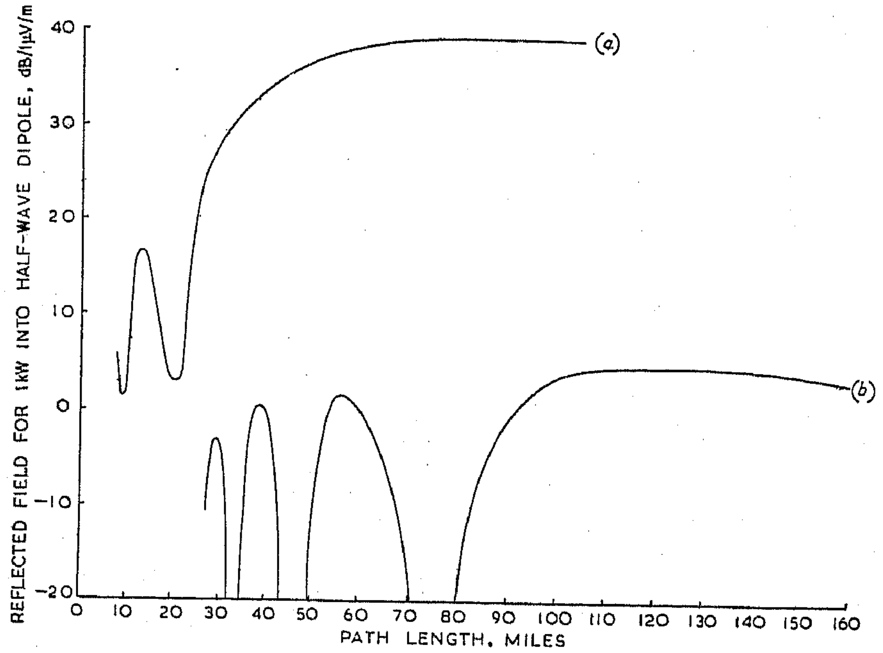


Fig. 23.—Theoretical tropospherically-reflected field strengths on 75 Mc/s.

(*a*) Pronounced layer.
Height, 2 000 ft.
Thickness, 160 ft.
$\Delta n = 2 \times 10^{-5}$.

(*b*) Average layer in the United Kingdom.[11]
Height, 4 500 ft.
Thickness, 500 ft.
$\Delta n = 5 \times 10^{-6}$.

Fig. 23 shows, for terminal heights much less than the layer height, the theoretical tropospherically reflected field-strength on 75 Mc/s for (*a*) a fairly pronounced elevated layer, and (*b*) an average layer in the United Kingdom. The curves are plotted for 1 kW radiated power from a half-wave dipole, and are derived from charts of tropospheric reflection coefficients prepared by Millington.[16] The average layer characteristics [Fig. 23(*b*)] are taken from meteorological data collated by Saxton.[11] The curves are based on a single reflection at the layer, and extend to a range corresponding to the horizon for the reflection point on the layer. Near this extreme range, earth diffraction effects and contributions from other tropospheric ray paths begin to become important, and the latter influence is likely to sustain the orthodox field well beyond the range of the curves. It is concluded that, except for short ranges, condition (*c*) may be approximately fulfilled, particularly for well-pronounced layers fairly near the ground.

Unless the reflecting layer is thin in terms of the wavelength (an unlikely contingency in the v.h.f. band), the layer reflection coefficient will, in general, be frequency-dependent, and therefore condition (*d*) cannot strictly hold. We can thus expect the suggested relationship to hold over only a limited frequency band.

It is concluded from the above reasoning that paths of varying lengths and topography which are subject to identical tropospheric layers may, if not too short, experience fades whose depth is uniquely related to their respective median path-attenuation. In practice all radio paths are, of course, subject to layers whose characteristics are varying, but the conjectured relationship between median path-attenuation and fading should still hold if the paths are all subject to statistically similar tropospheric variations. This last condition is difficult to verify meteorologically, and moreover the required degree of conformity is not easy to assess. It seems probable, however, that most of the observed departures from the suggested relationship are attributable to climatic differences to which the paths were subject during test (Section 3.4.1).

To simplify the argument we have up to now implicitly defined depth of fading as the difference between the median level and the level at the bottom of a fade. A more practical approach is to replace the lowest fading level by the level which was exceeded for some very large percentage of the time. From the engineering viewpoint, fading range is usefully defined as the difference between the median level and the level exceeded for 99·9% of the time. This definition will be adopted henceforth. It can equally well be written in the form "the difference between the median path-attenuation and the path attenuation exceeded for 0·1% of the time." For specified meteorological conditions, which are in any case a prerequisite of our conjectured relationship, we can interpret depth of fading in the foregoing way without impairing the validity of the relationship.

(3.4.1) Confirmatory Tests of the Conjectured Relationship.

To test the validity of our supposition we shall look for a connection between observed median path-attenuation and fading range (as defined above), computed for all paths for which a few days' continuous v.h.f. signal recordings are available. When taking these records it was impossible to wait for specific weather conditions, and, in fact, many varieties of climate were sampled in the course of many tests in scattered tropical regions. No attempt has been made to classify the types of weather experienced, and as the degree of tropospheric similarity demanded by the conjectured relationship is in any case obscure, it was decided to recognize initially only the well-established distinction between land and maritime meteorology, by sorting the radio paths into over-land and over-water categories. This broad distinction was clearly revealed by comparing mass plots of fading range against median path-attenuation for paths in each of these classes. The over-land plot revealed a further distinction between two types of land fading, which were often associated with particular countries. The over-water plot showed much greater unanimity of fading characteristics, even when a few Mediterranean paths were included in the analysis, which is otherwise confined to tropical regions.

Though most of the tests were carried out on about 170 Mc/s, a few measurements on about 80 Mc/s were made, in some cases in conjunction with the higher frequency. The results showed no significant difference between fading properties on these two frequency bands, and the few measurements on the lower band have therefore been merged with those on the higher band.

(3.4.2) Over-Land Fading Ranges.

Eighty-two tropical over-land paths have been analysed, involving path lengths between 12 and 106 miles and terminal heights ranging between 50 ft and 10 000 ft above sea level;

31 paths were optical and 51 paths were grazing or obstructed. Measurements were spread over some three years, and included 77 tests in the 170-Mc/s band and nine in the 80-Mc/s band. Each test comprised about six days' continuous-signal recording.

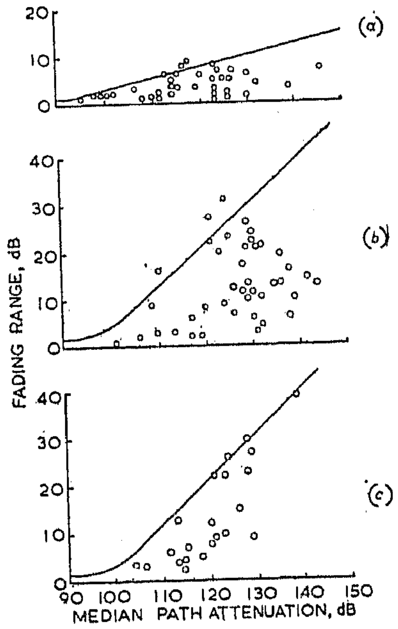


Fig. 24.—Mass plots of fading range versus median path-attenuation.

(*a*) Land paths in Malaya, Ceylon and East Africa.
(*b*) Land paths in Nigeria and the Gold Coast.
(*c*) Water paths in Malaya, the Gold Coast, East Africa and Mediterranean.

Fig. 24(*a*) is a mass plot of fading range against median path-attenuation for Malaya (26 paths), Ceylon (eight paths), and East Africa (five paths). Although correlation between the variables is not statistically significant, this classification has engineering value in that the scatter of points is confined within a reasonable compass. The above three regions show similar fading characteristics which are markedly less severe than for other tropical regions investigated. In the case of Malaya and Ceylon particularly, the small degree of fading results mainly from diurnal refraction variations [Fig. 17(*a*)], there being a striking absence of the deep and relatively rapid fading characteristic of atmospheric layering. The Malayan tests were sufficiently protracted to form strong evidence that the condition is a permanent feature of that country.

Fig. 24(*b*) is a mass plot for Nigeria (31 paths), and the Gold Coast (12 paths). Although significantly correlated, the points are very scattered, for the main reason that about half the examples for these two countries exhibited abnormally high fading ranges attributable to the influence of intense nocturnal surface ducts (Section 2.2.2).

(3.4.3) Over-Water Fading Ranges.

Seventeen predominantly over-water paths have been analysed, involving path lengths between 24 and 200 miles and terminal heights ranging between 100ft and 6 500ft above sea level. Nine paths were optical and eight were grazing or obstructed. Measurements were spread over some four years and included 18 tests in the 170-Mc/s band and five in the 80-Mc/s band.

Fig. 24(*c*) is a mass plot for Malaya (four paths), East Africa (four paths), the Gold Coast (five paths), and the Mediterranean Sea (four paths). In this case the points exhibit good correlation as well as small scatter, despite the large geographical separation of many of the paths.

(3.4.4) General Comments on Fig. 24.

Fig. 24 shows that the conjectured unique relationship between fading range and median path-attenuation is realized to a limited, but nevertheless useful, degree in the sense that the scattered points lie within fairly well-defined bounds. From the engineering viewpoint, the prediction of the likely maximum fading ranges to an accuracy no better than 10dB, say, has considerable value.

The representative lines on the charts are designed to give cautious estimates of fading ranges in terms of median path-attenuation. The following paragraph will make it clear that the fading range, as defined here, is likely to fluctuate more widely when measured over short periods than over long. The communication engineer would like an assurance that the fading range over a short period is unlikely to exceed a specified value. For this reason, the representative lines have been placed to give a strong probability of conservative fading-range estimates on a weekly basis (it being remembered that the experimental points are derived from tests of about this duration).

The fading range, as computed from only a few days' record, is very sensitive to the incidence of even a single deep fade, owing to the fact that, by definition, it is based on signal levels exceeded for all but 1½ minutes a day (99·9% of the time). In this way, the exclusion, for example, of the last hour of a 150-hour record could materially affect the computed fading range. Some of the scattering of the points on Fig. 24 is attributable to this cause, particularly, of course, under conditions of deep fading such as occurred nocturnally in Nigeria and the Gold Coast. The scatter of points could be reduced, either by computing fading range on a 99% basis, or by recording for longer periods. The latter expedient is usually impracticable, and it has seemed best to the author to retain the original definition of fading range (99·9% basis) because of its engineering value, and to accept the increased scatter of the points.

The following further conclusions (expected to apply over the frequency range 80–200Mc/s), emerge from a study of Fig. 24:

(*a*) Fading is negligible for all v.h.f. paths having median attenuations less than about 100dB.

(*b*) For path attenuations exceeding about 100dB, the trend of the points suggests an approximately linear increase of fading range with median path-attenuation.

(*c*) With the exception of those paths in Nigeria and the Gold Coast subject to nocturnal surface ducts, land paths fade less than water paths having the same median attenuation. This confirms that atmospheric stratification is, in general, less marked over land than over water.

(*d*) The good correlation for water paths is attributable to their simple geometry coupled with the likelihood of regular atmospheric stratification, which together favour the conditions justifying the assumptions of Section 3.4, on which the expected relation is based.

(*e*) Over-land paths in Malaya, Ceylon and East Africa fade less on the average than over-land paths in Nigeria and the Gold Coast having the same median attenuation.

(*f*) For a given median path-attenuation, those over-land paths in Nigeria and the Gold Coast subject to nocturnal surface ducts fade about as much as over-water paths.

(*g*) The scattering of points masks any difference there may be in the fading characteristics on 80Mc/s and 170Mc/s.

(3.4.5) Use of the Fading-Range Charts.

In specifying the performance of radio links it is often laid down that a certain signal/noise ratio shall be exceeded for 99·9% of the time. For a specified equipment this places an upper limit on the path attenuation which can be exceeded for 0·1% of the time.

When planning v.h.f. communications in the regions discussed, we can use the charts of Fig. 24 to ensure that the system approaches the desired performance. For suitably situated routes, all that is required is an estimate of the median path-attenuation, from which the associated fading range is conservatively estimated by reference to the appropriate chart in Fig. 24. By definition, the sum of the fading range and the median path-attenuation gives the path attenuation likely to be exceeded for 0·1% of the time, which is the required crucial property of the path.

The median path-attenuation may be estimated, either by a short signal-strength check (which need not be long enough to reveal the fading range accurately), or by computation where the profile allows of this. In the latter event the median path-attenuation should be based on the expected average refraction ($k = 4/3$, for example). An example of the last method is given in Section 3.3.1, but in that case an exceptional profusion of Mediterranean Sea data permitted a more comprehensive estimate of statistical path behaviour than is at present possible elsewhere.

(3.4.6) Reliability of the Fading-Range Charts.

It is clear from Fig. 24 that v.h.f. fading-range forecasts may now be possible to a useful accuracy for land paths in certain specified regions, and for widely scattered water paths. Predictions for land paths in Nigeria and the Gold Coast are subject to large uncertainty, but should err mostly on the conservative side if the foregoing recommendations are followed.

Fading behaviour for regions outside the scope of the present measurements cannot be forecast with confidence, but the use of Fig. 24(*a*) is provisionally suggested for over-land tropical and sub-tropical paths where there is no meteorological evidence of the marked nocturnal surface ducts responsible for the excessive fading evident from Fig. 24(*b*). Fig. 24(*c*) is provisionally recommended for predominantly water paths in tropical regions at all seasons, and for sub-tropical and Mediterranean Sea areas in summer. In the two latter regions, winter fading ranges are probably somewhat less than predicted by Fig. 24(*c*). Moreover, it is likely that paths in temperate and arctic regions will generally exhibit smaller fading ranges than forecast by the charts.

(4) CONCLUSIONS

An analysis of continuous v.h.f. signal-strength recordings over numerous and diverse paths in tropical and Mediterranean regions has revealed signal variations which are satisfactorily explainable in terms of well-known tropospheric states. Furthermore, it has been possible in some cases to reconstruct from the signal records the salient features of tropospheric *M*-profiles operative during the tests.

Prolonged measurements over a Mediterranean radio-path have yielded useful statistics of variations in the effective earth-radius factor, which should assist in predicting the statistical behaviour of other radio paths in that area.

A limited connection has been found between median path-attenuation and fading range for paths experiencing comparable tropospheric influences, which, with various reservations, makes possible the approximate forecasting of the statistical behaviour of radio paths without assistance from a preliminary radio survey.

(5) ACKNOWLEDGMENTS

The author is indebted to Marconi's Wireless Telegraph Co., Ltd., for permission to publish the paper. He also wishes to thank the following Administrations for permission to use material obtained during radio surveys in their territories:

The Malaya Telecommunications Department.
The Ceylon Department of Posts and Telecommunications.
The East Africa Posts and Telegraphs Department.
The Nigeria Posts and Telegraphs Department.
The Posts and Telecommunications Department of the Gold Coast.
Cable and Wireless Ltd.
The Ministry of Transport and Communications, Israel.

The author would also like to thank his colleagues, Messrs. G. Millington and G. A. Isted, for many helpful suggestions in the preparation of the paper.

(6) REFERENCES

(1) Schelleng, J. C., Burrows, C. R., and Ferrell, E. B.: "Ultra-Short-Wave Propagation," *Proceedings of the Institute of Radio Engineers*, 1933, **21**, p. 427.

(2) Eckersley, T. L., and Millington, G.: "Application of the Phase Integral Method to the Analysis of the Diffraction and Refraction of Wireless Waves round the Earth," *Philosophical Transactions of the Royal Society*, A, 1937, **237**, p. 273.

(3) Domb, C., and Pryce, M. H. L.: "The Calculation of Field Strengths over a Spherical Earth," *Journal I.E.E.*, 1947, **94**, Part III, p. 325.

(4) Englund, C. R., Crawford, A. B., and Mumford, W. W.: "Ultra-Short-Wave Transmission and Atmospheric Irregularities," *Bell System Technical Journal*, 1938, **17**, p. 489.

(5) Booker, H. G.: "Elements of Radio Meteorology. How Weather and Climate cause Unorthodox Radar Vision beyond the Geometrical Horizon," *Journal I.E.E.*, 1946, **93**, Part IIIA, p. 69.

(6) Booker, H. G., and Gordon, W. E.: "A Theory of Radio Scattering in the Troposphere," *Proceedings of the Institute of Radio Engineers*, 1950, **38**, p. 401.

(7) Megaw, E. C. S.: "Scattering of Electromagnetic Waves by Atmospheric Turbulence," *Nature*, 1950, **166**, p. 1100.

(8) Appleton, E. V.: "The Influence of Tropospheric Conditions on Ultra-Short-Wave Propagation," Meteorological Factors in Radio Wave Propagation (Physical Society, London, 1946), p. 1.

(9) Sheppard, P. A.: "The Structure and Refractive Index of the Lower Atmosphere," *ibid.*, p. 37.

(10) Durst, C. S.: "Radio Climatology," *ibid.*, p. 193.

(11) Saxton, J. A.: "The Propagation of Metre Waves beyond the Normal Horizon," *Proceedings I.E.E.*, Paper No. 1112 R, March 1951 (**98**, Part III, p. 360).

(12) "Radio Wave Propagation" (Academic Press, New York, 1949), Vol. II, p. 223.

(13) Bussey, H. E., and Birnbaum, G.: "Measurement of Variations in Atmospheric Refractive Index with an Airborne Microwave Refractometer," *Journal of Research of the National Bureau of Standards*, 1953, **51**, No. 4, p. 171.

(14) Megaw, E. C. S.: "Waves and Fluctuations," *Proceedings I.E.E.*, 1953, **100**, Part III, p. 1.

(15) Epstein, P. S.: "Reflection of Waves in an Inhomogeneous Absorbing Medium," *Proceedings of the National Academy of Sciences*, 1930, **16**, p. 627.

(16) Millington, G.: "The Reflection Coefficient of a Linearly Graded Layer," *Marconi Review*, 1949, **12**, p. 140.

(17) Booker, H. G., and Walkinshaw, W.: "The Mode Theory of Tropospheric Refraction and its Relation to Wave-Guides and Diffraction," Meteorological Factors in Radio Wave Propagation (Physical Society, London, 1946), p. 80.

(18) Day, J. P., and Trolese, L. G.: "Propagation of Short Radio Waves over Desert Terrain," *Proceedings of the Institute of Radio Engineers*, 1950, **38**, p. 165.

(19) Booker, H. G.: "A Rough Sketch of World Radio Climatology over Sea," T.R.E. Report No. 1730, 1944.

(20) "Radio Wave Propagation" (Academic Press, New York, 1949), pp. 21–24, 200–205, 223–225, 251–256.

621.315.612.4 : 621.38

Paper No. 1752 M
Jan. 1955

NOISE GENERATION IN CRYSTALS AND IN CERAMIC FORMS OF BARIUM TITANATE WHEN SUBJECTED TO ELECTRIC STRESS

By A. C. KIBBLEWHITE, M.Sc., Ph.D.

(*The paper was first received 7th January, and in revised form 1st July,* 1954.)

SUMMARY

The paper describes an investigation of the noise generated in crystals and in ceramic forms of barium titanate when subjected to electric stress, as a function of the stress and of temperature above and below the Curie point of the material. The primary purpose of the work was to establish the existence of a ferro-electric Barkhausen effect associated with the domain structure. The results obtained on crystals have permitted an estimate of the effective volume of the individual domains, and this is compared with a previously published value. The ceramic specimens exhibited considerable additional noise of different origin. The characteristics of this noise have been studied, but its satisfactory elucidation will require much further investigation.

Part 1.—MEASUREMENTS ON BARIUM TITANATE CRYSTALS

INTRODUCTION

Considerable literature now exists concerning the electrical properties of materials based on barium titanate and the marked similarity between these properties and the magnetic characteristics of ferromagnetic materials. For instance, their permittivity and the permeability of ferromagnetic materials are similar complex functions of field strength and temperature within a temperature range bounded by Curie points; they exhibit permanent dielectric polarization and dielectric hysteresis effects which correspond to remanent magnetization and magnetic hysteresis; and magnetostriction has its counterpart in an electrostrictive behaviour in which the mechanical strain is similarly proportional to the square of the applied field-strength.

The cause of these similarities resides in the fact that both types of material exhibit a domain structure at temperatures below their respective Curie points. For pure barium titanate this point occurs at 120°C, where the crystal symmetry changes with decreasing temperature from cubic to tetragonal form. In ferromagnetic materials the presence of domains is only revealed directly by careful surface studies on suitably prepared crystals,[1,2] but it is readily seen under the polarizing microscope in thin flake-like barium titanate crystals grown from the melt. In consequence, it has been possible to observe the way in which the domain structure of such crystals is influenced by electrical and mechanical stress and by temperature,[3,4] to a degree not yet achieved with ferromagnetic materials.

One of the experiments designed to test the validity of the original Weiss domain theory of ferromagnetic behaviour was the detection of what is now known as the Barkhausen effect. It was agreed that if, as postulated by the domain theory, the magnetization process were discontinuous, the discontinuities should be audible as noise pulses in a suitable detecting circuit. The occurrence of such "clicks" or noise was first demonstrated by Barkhausen,[5] and it seems reasonable to expect a corresponding phenomenon in barium titanate under conditions of varying applied electric-stress.

(1) PREVIOUS EVIDENCE FOR AND AGAINST A BARKHAUSEN EFFECT IN BARIUM TITANATE

Several attempts have been made to demonstrate the Barkhausen effect in barium titanate,[6–10] but analysis showed them to be generally unsatisfactory, essentially qualitative and varying in their conclusions from confirmation of the phenomenon to complete denial of its existence. The reason for the ambiguities was that, in contrast to ferromagnetic materials, there are several possible mechanisms of noise generation in dielectric materials subjected to high electric stress. They are: (*a*) noise associated with ionic conduction through the volume and over the surface of the specimen; (*b*) noise resulting from imperfectly applied electrodes—thus if these consist of thin metallic films containing discontinuities, discharges may occur between isolated elements and the main film; (*c*) noise associated with discharges in the ambient medium or in microvoids within the specimen; and (*d*) noise of electronic origin preceding internal breakdown of the material.

It has been found possible with single crystals to avoid these complicating factors, and it is significant that the most satisfactory previous investigation is that on a single crystal by Newton, Ahearn and McKay.[8]

(2) NATURE OF THE BARIUM TITANATE CRYSTALS STUDIED

A number of crystals of surface area a few square millimetres and thickness a few tenths of a millimetre were examined under the polarizing microscope at room temperature, and exhibited patterns of the nature shown in Fig. 1. The crystals were thoroughly cleaned in acid, washed in alcohol and dried before an attempt was made to attach electrodes. Initially, evaporated electrodes of silver and gold were attached, but these proved

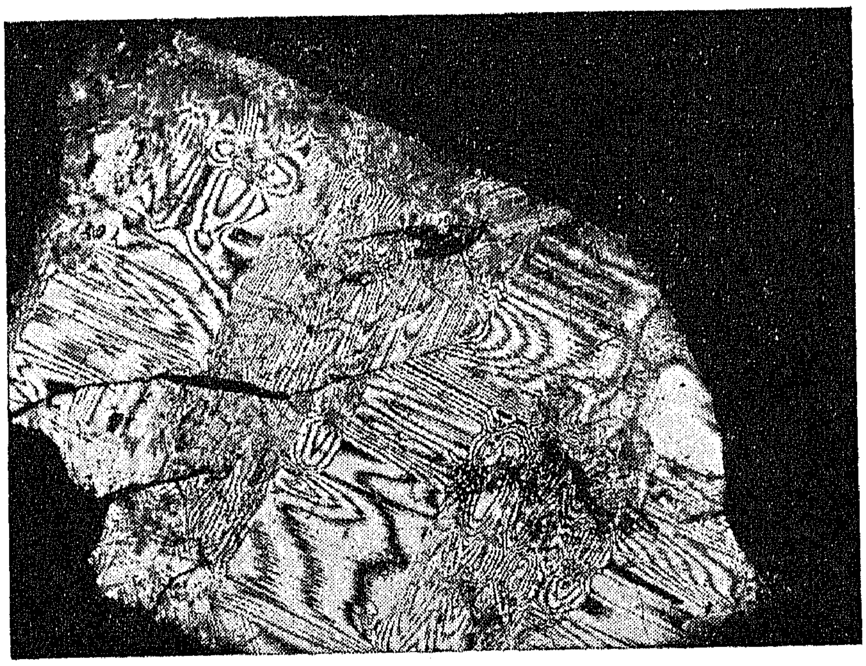


Fig. 1.—Crystal photographed between crossed Nicol prisms in monochromatic light (×30).

Written contributions on papers published without being read at meetings are invited for consideration with a view to publication.

Dr. Kibblewhite was formerly in the Electrical Engineering Department, Imperial College of Science and Technology, University of London.

insufficiently stable and robust, and use was finally made of silver-paste electrodes fired at about 500°C.

(3) EXPERIMENTAL ARRANGEMENTS

The main item of the experimental assembly consisted of a 5-stage selective amplifier covering the frequency range 200c/s to 100kc/s which incorporated certain special features to reduce the inherent noise level to a minimum. At the frequency setting of 4·12kc/s with a bandwidth of 1kc/s, used throughout the investigation to be described, the equivalent input noise signal was determined mainly by the thermal noise of the input circuit and was somewhat less than one microvolt. The gain available at this amplifier setting was 120dB.

The noise signal after amplification was fed into an "infinite impedance" detector and the level read on a measuring instrument in the plate circuit. In those measurements involving continuous recording, the amplified signal before rectification was fed into a Brüel and Kjaer level recorder, and the noise records shown later were taken in this way. In addition to these means of indication, the noise fluctuations could be displayed on an oscillograph, so that any extraneous disturbances were immediately evident and measurements could be suspended until these were removed.

The input circuit used is shown in Fig. 2. R_1 and C_1 constitute the charging circuit controlling the exponential rate of rise of stress across the specimen C_s; this could be varied by changing R_1. C_2 was included as a safety measure and to ensure against an alteration in the operating conditions of the first amplifier stage in those cases where C_s was conducting appreciably. The capacitance of both C_1 and C_2 was 1 μF and, for reasons mentioned in Section 4.3.1, R_s was fixed at 1 megohm. An analysis of the thermal-noise input voltage of this circuit shows that it increases as the capacitance of C_s decreases; the calculated and measured values are compared in Fig. 3.

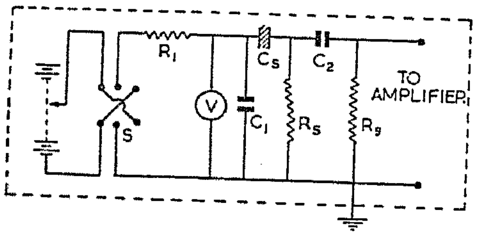


Fig. 2.—Input circuit.

The direct-voltage supply consisted of a bank of dry cells mounted on blocks of paraffin wax and completely shielded in a metal box.

The procedure adopted was to charge the specimen capacitor to the required stress and record the noise output so produced. The rate of charging could be altered at will and the stress reversed by switch S. An electrostatic voltmeter, V, indicated the voltage across the specimen.

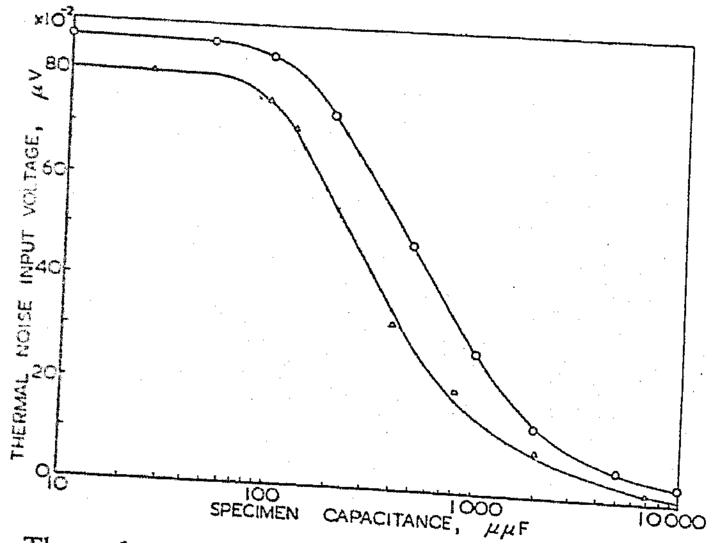


Fig. 3.—Thermal-noise voltage of input circuit as a function of specimen capacitance.

○ ○ Calculated noise.
△ △ Measured noise.
Temperature 20°C.

(4) STATEMENT OF RESULTS

Considerable electrical noise was observed when a crystal was connected in the input circuit of the noise measuring equipment and subjected to an electrical stress. Since no noise was observed when the crystal was replaced by a good non-ferroelectric capacitor, this noise was definitely characteristic of the crystal. The noise was impulsive in form and died away soon after the stress had reached its full value. Nearly all the pulses were in the same direction as viewed on the oscillograph and reversed their direction when the polarity of the applied stress was reversed. In an attempt to identify the source of the noise, its behaviour with regard to certain variable parameters was investigated as detailed in the succeeding Sections. In all cases, the overall gain of the equipment was so adjusted that a full-scale deflection of the level recorder corresponded to a noise input signal 50dB above the basic thermal-noise level of about one microvolt (see Fig. 3).

(4.1) Noise Output as a Function of Electrical Stress

At field strengths less than about 400 volts/cm no noise was observable, but with further increase of stress large noise pulses began to appear in considerable numbers. This noise was not confined to the time during which the stress was building up exponentially, but was found to continue for some seconds after the stress had reached its full value. Further increases of stress produced additional noise, with the noise resulting from each increment becoming smaller and smaller beyond 5kV/cm. On reducing the stress to zero, the pulses constituting the noise were predominantly of the reverse polarity. If the stress was reduced in small steps the greatest noise activity occurred as the voltage approached zero.

Reversing the field produced a similar noise response. In most of the crystals investigated the noise outputs in the two field directions were significantly different in magnitude. An interchange of the responses when the specimen was reversed showed that this was not a feature of the measuring equipment. In those crystals in which the effect was particularly pronounced, it was noted that the hysteresis loops were also slightly asymmetrical, this asymmetry no doubt resulting from crystal strains.

If, after stress removal, the same maximum stress was reapplied in the same direction, the noise output was found to be much smaller than on the first application. This dependence on the previous electrical history of the specimen seemed, however, to be confined to conditions of increasing field, since, when the field was again reduced to zero, the "reverse" or "discharge" noise was usually as great as on the first discharge. This reverse

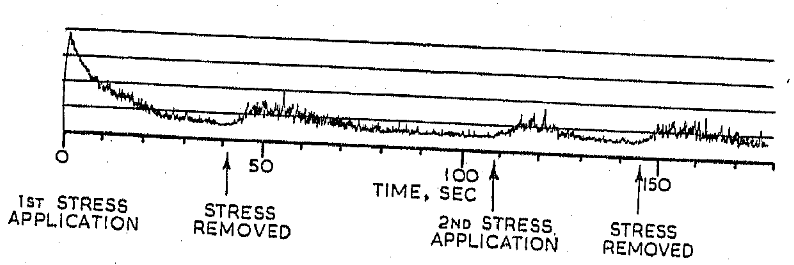


Fig. 4.—Noise response to the application and removal of a stress of 2·5kV/cm.

noise seemed to become more pronounced the longer the field was left applied. Some of the above effects are illustrated in Fig. 4.

(4.2) Noise Response as a Function of Temperature

(4.2.1) Under Electrical Stress.

Before attempting to measure the noise output as a function of temperature, the specimen was suitably conditioned by subjecting it to a series of charge–discharge cycles at the stress to be employed. For the purposes of this investigation, the stress was kept reasonably low, 2·5kV/cm, to reduce as much as possible the complication occasioned by conductivity at temperatures around the Curie point.

The noise output was then recorded at various temperatures up to and above the Curie point at 120°C. A sequence of stress-reversal records is shown in Fig. 5. By considering the noise outputs for, say, the negative cycle of the applied stress and measuring the areas under the envelope with a planimeter, the same time period being involved in each case, the total noise-output/temperature plot of Fig. 6(*a*) was obtained.

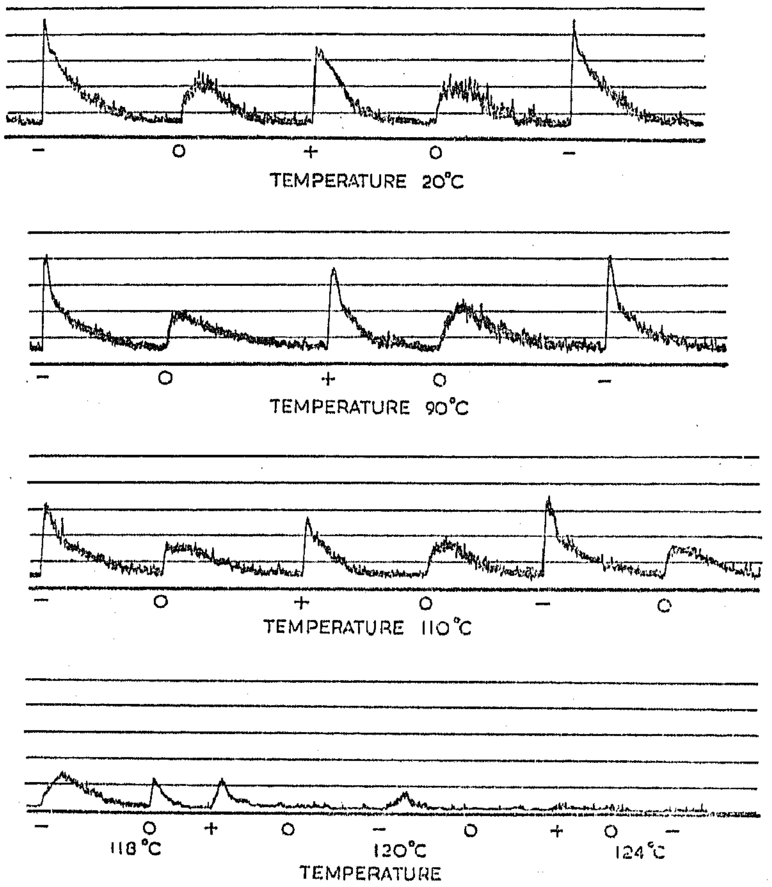


Fig. 5.—Typical noise records at various temperatures for an applied stress of 2·5kV/cm.

It will be seen that the noise falls to zero at a temperature close to the Curie point. At temperatures about 10°C above this point noise reappeared; its level increased rapidly with temperature and stress, and showed no sign of decaying with time. This was readily identified with a sharp increase in the conductivity of the crystal, which became marked at these temperatures.

(4.2.2) Stress-Free Response.

It was observed that considerable noise outputs were recorded as the specimen was heated and cooled, even though no electrical stress was applied. Fig. 7(*a*) shows a typical noise curve as the crystal was cooled through the Curie point, and Fig. 7(*b*) the behaviour with a stress of 12·5kV/cm applied during the cooling.

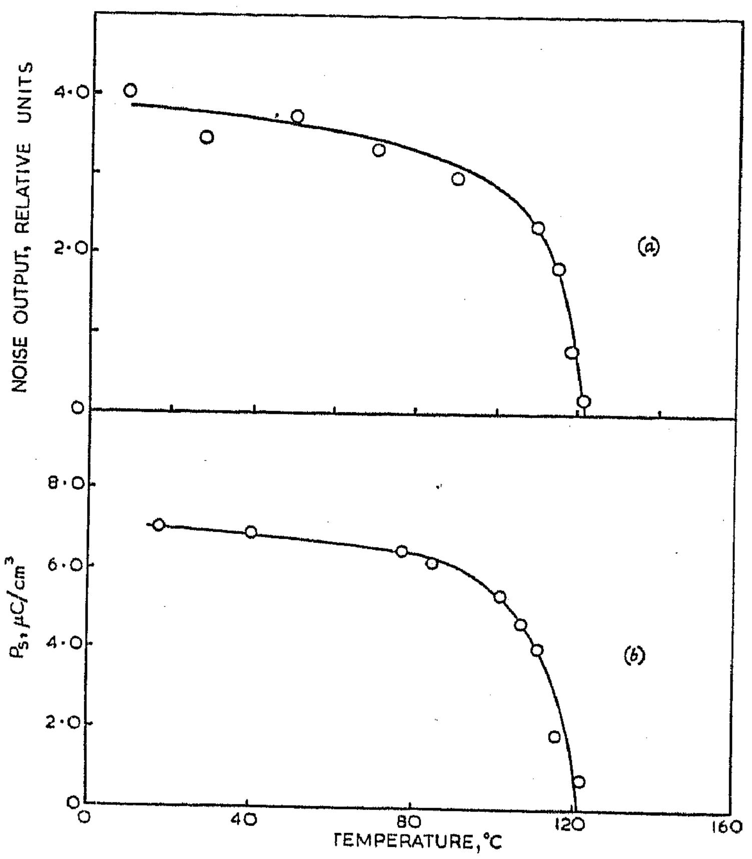


Fig. 6.—Noise response and spontaneous polarization plotted against temperature.

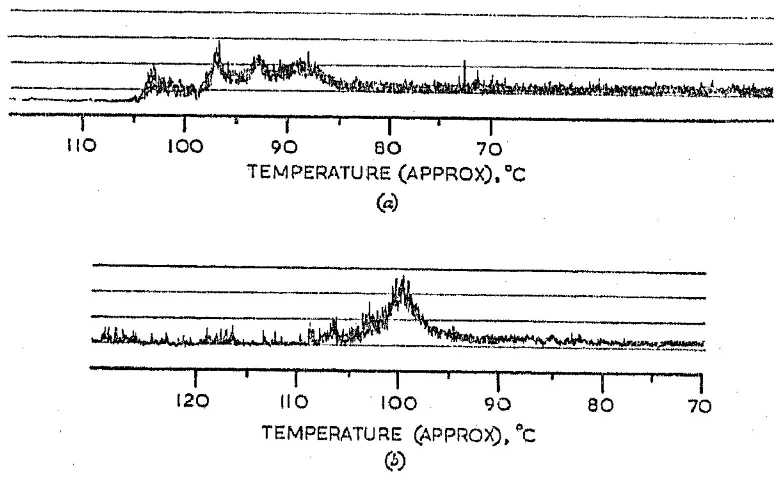


Fig. 7.—Noise record during cooling through the Curie point without and with an applied polarizing stress.

(*a*) Cooling free of stress.
(*b*) Applied d.c. stress. 12·5kV/cm.

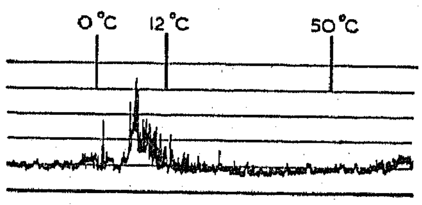


Fig. 8.—Noise record during heating from below 0°C without polarizing electric stress.

Fig. 8 shows a noise record taken as the crystal was heated through the transition point which occurs in barium titanate around 5°C.

(4.3) The Shape, Size and Duration of the Noise Pulses

With a view to obtaining information on the process from which the pulses arose, a closer study of the form of the pulses was undertaken. Before any reliable conclusions could be reached it was essential to know to what extent the method of detection distorted the initial discontinuity. In these investigations the amplifier was used in a broad-band condition with a bandwidth of some 2 Mc/s. Independent check measurements showed that this bandwidth was sufficient to ensure a faithful reproduction of the pulses encountered.

(4.3.1) The Influence of the Input Circuit.

Most of the pulses observed were of simple form, but an occasional one exhibited a double-hump structure. Investigation revealed that the trailing edge of the pulses was determined essentially by the time-constant C_sR_s of the input circuit (see Fig. 2). The leading edge and amplitude, on the other hand, were unaffected by C_sR_s provided that this time-constant was appreciably larger than the rise time of the pulse. A value of C_sR_s around 10^{-4} sec was found to be sufficiently large, in general, to ensure this. The rise time was dominantly of the order of 10^{-5} sec but ranged upwards from this figure to 10^{-4} sec and beyond. A similar distribution occurred in the pulse amplitude, the outside limits of which appeared to be 7 and 35 microvolts, with the maximum of the distribution occurring at 12 microvolts.

(5) DISCUSSION OF RESULTS

(5.1) The Origin of the Noise

Optical investigation of the crystals confirmed the well-known fact that the domain structure shown in Fig. 1 was influenced by mechanical and electrical stress. The changes brought about were never smooth but took place jerkily. It seemed reasonable to associate the noise with these structural changes, and in order to facilitate the discussion of the foregoing results, it is assumed that the noise observed is really the ferro-electric analogue of the Barkhausen effect of ferromagnetic materials. Each piece of evidence will then be analysed with a view to determining the support, or otherwise, that it lends to this hypothesis.

(5.1.1) The Noise/Temperature Relationship.

Figs. 5 and 6(*a*) show that the noise produced by an applied field falls to zero at about 120° C, the temperature at which the domain structure of barium titanate disappears. This correspondence is a necessary condition for the noise to be Barkhausen in origin. Furthermore, since the Barkhausen effect is associated with the "irreversible" increments in the total polarization, it should be intimately related to the spontaneous polarization. The striking nature of the relationship between the spontaneous moment, P_s (as calculated from 50 c/s hysteresis loops) and the noise response can be seen in Fig. 6(*b*). This result alone is strong evidence that the noise observed is the external manifestation of a Barkhausen effect. The electrical noise which resulted when the crystal was heated or cooled free of electrical stress, Fig. 7(*a*), supports this conclusion. It is seen that the noise is greatest over a temperature range some 20–30° C from the Curie point, over which the spontaneous polarization builds up to its full value. Although the domain structure appeared stable by the time the temperature had fallen to 90° C, the noise records reveal that minor readjustment continues down to room temperature.

It was found that the noise records of crystals displaying a square-net domain pattern were always the same in form irrespective of the number of times the heating and cooling cycle was repeated, whereas this was not so with other crystals. It is of interest to note, therefore, that according to Forsbergh,[11] only the former type of crystal reverts to the same structure on cooling. Forsbergh accounts for this reproducibility of domain structure in terms of crystal strains which predetermine the final pattern even at temperatures above the Curie point. The reproducibility of the noise records in certain crystals would appear to provide additional confirmation, therefore, of the connection between the noise and the domain structure. This is strengthened by the noise around 5° C—shown in Fig. 8—an occurrence which would seem to be associated unquestionably with the first of the two lower transition points of barium titanate.

(5.2) The Noise as a Function of Stress

The majority of the noise/stress characteristics are also readily interpreted on the basis of the Barkhausen effect.

If the polarization in a ferro-electric material increases in discontinuous "jumps" under the influence of an applied field, the Barkhausen noise must be impulsive in form. The polarity of these pulses should reverse when the field is reversed; and those appearing when the specimen is discharged, owing to the re-randomization of domain vectors, should also be of opposite polarity to those appearing when the field is applied. Because of remanence the polarization does not decay appreciably, so the noise output should depend appreciably on the previous electrical history of the specimen. Reference to Section 4.1 shows that the observed noise possessed all these characteristics.

(5.3) The Effective Domain Volume

The phenomenon of Barkhausen noise is the external manifestation of the alignment of domains into the direction of an applied field. From a knowledge of the size of the individual pulses it is possible to calculate a value for the volume of the ferro-electric material within which the polarization changes discontinuously.* The calculation is very similar to that described by Newton, Ahearn and McKay,[8] and results in an expression for this "effective" volume

$$v = \frac{4 \cdot 43 dV \epsilon_r A}{P_s} 10^{-10} \mathrm{cm}^3$$

where dV = Pulse amplitude, volts.
ϵ_r = Relative permittivity.
A = Area of electrodes, cm².
and P_s = Spontaneous polarization, coulombs/cm².

Substitution of appropriate values for the crystals examined gives a value, $v = 10^{-8}$ cm³, which is larger than the value of 10^{-9} cm³ obtained by Newton *et al.*

This calculation is of little significance, however, since it is impossible to decide the value to be given to the relative permittivity. Newton *et al.* overlooked this fact, but the stress dependence of ϵ_r is such that any value from 2 000 to 20 000 might be permissible. In the present case this would produce equivalent variations in v from 10^{-7} to 10^{-8} cm³. Moreover, the value taken for the pulse amplitude is an average one; measurements revealed a distribution in pulse amplitudes from 7 to 35 microvolts.

It is possible, however, to obtain an estimate of v which does not depend on either of the doubtful quantities, ϵ_r and dV. This is based upon the total number of pulses recorded during a given application of stress.

Although limitations in the counter employed did not allow a direct count of the total number of pulses produced by a field, which would have ensured effective saturation (say, 12 kV/cm), a fairly reliable estimate of this number was obtained as follows:

Fig. 9 indicates the pulse number recorded as a function of field strength, and was obtained using a reduced amplification

---

* This volume will not necessarily be the same as that of the optically visible domains seen in Fig. 1. The Barkhausen domain is probably more accurately identified as the volume swept out by a boundary in a single discontinuous jump over a potential-energy maximum. It is possible that several such discontinuities would occur in aligning each of the visible regions of Fig. 1.

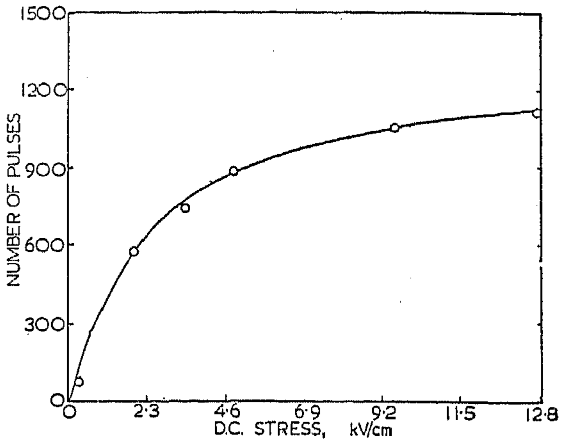


Fig. 9.—Number of pulses greater than a certain size plotted against applied d.c. field: the counting time is one minute in each case.

to avoid jamming of the counter and subsequent miscounting. It therefore refers only to those pulses greater than a certain minimum amplitude. To reduce errors due to time effects, counting was continued for one minute in each case. It indicates, however, that the polarization (number of pulses) produced by a stress of 2·5kV/cm is nearly one-half the number produced by the saturating stress of 12kV/cm, a fact which is also confirmed by the 50c/s hysteresis loops. Thus, by measuring the total pulse count produced by an applied stress of 2·5kV/cm (where jamming was not serious), an estimate of the number of pulses likely to be produced by a saturation field was obtained by doubling this number. This resulted in a total count of 6 000 pulses for the particular crystal concerned.

It is reasonable to assume that the application of such a high stress will effectively reverse the polarization throughout the whole volume of the crystal. A value for the average volume of the specimen in which the polarization is reversed discontinuously may then be obtained, by dividing the total volume of the specimen by the number of pulses. This results in an average value of $v = 5 \times 10^{-8}\mathrm{cm}^3$, which agrees well with that obtained previously.

It is felt that this value is more reliable than the one obtained by Newton *et al.*, for, in spite of the fact that they obtained a pulse count at 30kV/cm by a linear extrapolation of the number/stress relation obtained over a stress range 0–2kV/cm (which Fig. 9 shows is not justified), they still only obtain a total volume change due to polarization of some 6% of the specimen volume. Their claim that this is in close agreement with that of 7% deduced by Mason is of little import, since, in the case referred to, Mason was working with polycrystalline specimens where the degree of polarization at stresses adequate to saturate a single crystal is known to be very low. It must be admitted, however, that there is no reason to expect that the effective domain volume will be the same in two different crystals. The query placed on Newton's value of $10^{-9}\mathrm{cm}^3$ is not therefore prompted by its contradiction to that found in the present work, but rather in the argument outlined in the preceding paragraph.

(5.4) Other Aspects of Newton's Paper

The results of the present work agree qualitatively with those obtained by Newton *et al.* Certain discrepancies do, however, exist, some of which have been mentioned above. Others are worthy of mention. They refer to the necessity of placing the crystal in a vacuum before the Barkhausen effect could be observed satisfactorily. This was not found necessary in the present work, and it seems likely they were being troubled with surface conduction, which could readily account for the large scatter in their pulse-number/stress data.

The average pulse duration of 10^{-5}sec observed in the present work is longer than the 10^{-6}sec figure mentioned by Newton. However, pulse-duration data obtained from direct measurement are not particularly reliable, and the discrepancy may easily be an experimental one.

If we assume a linear dimension of $3\sqrt{v}$ for a domain and a pulse duration of 10^{-5}sec, a velocity of propagation of 370cm/sec is obtained, compared with Newton's value of 1 000cm/sec.

(5.5) Anomalous Behaviour of the Crystals

(5.5.1) Time Effects.

On the basis of the foregoing evidence it is concluded that the noise arises from the Barkhausen effect. If this is correct, those pulses persisting after the field has become steady can only be interpreted as additional increases in the polarization. In other words, the total polarization produced by a given field is not only a function of its magnitude but also of the time it is applied to the specimen. Visual observations confirmed that minor adjustments in the domain pattern took place for considerable times after the applied field had reached its full value. Moreover, measurements of the incremental permittivity revealed a small drift with time (see Fig. 10), the time factor involved being very much the same as that observed in the noise measurements (cf. Fig. 4).

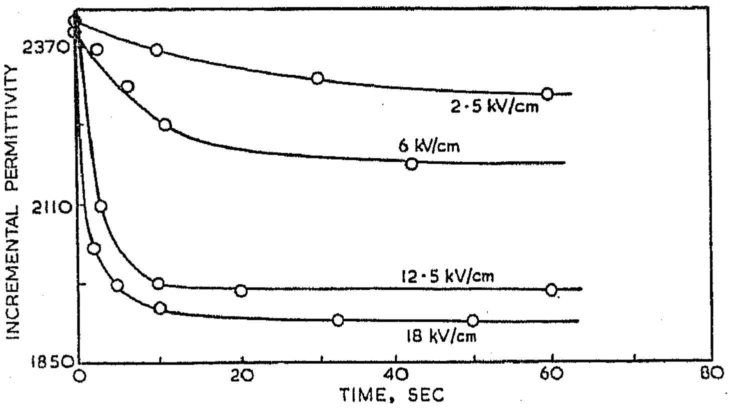


Fig. 10.—Variation of incremental permittivity with time under d.c. fields.

(5.5.2) Large Discharge Noise.

The only experimental result which does not fit conveniently into the general picture of the Barkhausen effect is that of the large noise response when the specimen is discharged (see Fig. 4). Because of remanence, which is demonstrated by the marked history dependence of the "forward" or "charging" noise, the large discharge noise cannot be accounted for satisfactorily by the re-randomization of domain vectors. This implies, of course, that it is not truly Barkhausen noise.

A study of a considerable number of records revealed that the discharge noise possesses the following significant characteristics: (*a*) the pulses of which it is composed are larger than those associated with the charging cycle; (*b*) there is usually a perceptible lag between the time the stress begins to decrease and the first appearance of this noise; (*c*) the maximum activity occurs at times subsequent to the stress reaching zero; (*d*) if the stress is applied in the same direction a second time, little noise is produced, but a considerable amount again appears on the discharge; and (*e*) the discharge noise seems to increase with the time of application of the polarizing field and also increases markedly with the magnitude of this polarizing field.

No satisfactory explanation has been found for this behaviour and the only proposal which appears at all probable is that the noise results from a type of piezo-electric effect associated with a mechanical relaxation of crystal strains. Some small measure of support for this hypothesis may be provided by the fact that pulses of comparable size occur also when the crystal passes through the 5° C transition point (see Fig. 8). Here, too, mechanical effects are pronounced. Moreover, the mechanism producing this noise is obviously a slow one, in view of the "lag" mentioned above. It is thus of interest to note that Bradfield[12] has reported that time variations in piezo-electric constants (which are proportional to crystal strain) seem to proceed at about one sixth of the rate of the time changes associated with the incremental permittivity (which is related to the polarization). There may be some justification, therefore, in assuming that the more sluggish components of the noise are associated with the purely mechanical effects of the polarization processes.

(6) CONCLUSIONS

It has been shown that when all the extraneous sources of noise that are not characteristic of the material itself have been removed, the noise displayed by crystals of barium titanate is essentially Barkhausen in origin. In view of Section 5.5.2, however, it appears that there is also a piezo-electric contribution to this noise arising from the purely mechanical effects of polarization. Since this is still associated with the domain structure of the material, it is justifiable to include it in the general term "Barkhausen effect." The occasional "double hump" pulse observed in this work and also mentioned by Newton *et al.*, probably arises from this contribution, when an initially stable region is triggered by the mechanical wave generated by the alignment of a neighbouring domain.

The general qualitative behaviour of the noise with such parameters as temperature is no doubt common to all crystals of barium titanate. It is most unlikely, however, that any quantitative generalizations can be made from determinations on the small selection of crystals which has been available for study. For the present case, this must apply to the calculation of an effective domain volume and estimates of pulse rise-times. A much more extensive investigation is needed to determine the influence of crystal size and shape on these quantities. The present results serve a useful preliminary purpose, but undoubtedly investigations on truly single-domain crystals and regularly shaped crystals would be more instructive.

It is felt that these would reveal a state of affairs similar to that which the modern view believes to exist in the equivalent ferromagnetic problem. It is now believed that there is no direct connection between the size of the observed Barkhausen pulses and the actual physical dimensions of the ferromagnetic domains, which appear to be very dependent on the crystal size and shape.[1,2] This is a complete reversal of the conclusions reached in earlier measurements on the Barkhausen effect.[13]

Part 2. MEASUREMENTS ON CERAMIC SPECIMENS

INTRODUCTION

Section 7 deals with randomly selected commercial capacitors incorporating barium–titanate-based ceramic material. It was found that excessive noise outputs in these capacitors could be traced to sources which were not characteristic of the material itself, and it was concluded that to obtain useful information on the latter would involve great care in the selection of the test specimens. Section 8 deals with the investigations carried out on a number of such specimens.

(7) RANDOMLY SELECTED COMMERCIAL CAPACITORS

The capacitors investigated included ones composed of pure forms of barium titanate and of its strontium and lead titanate variations, the specimens being either plate-like or tubular in shape and complete with fired-on electrodes.

When these were subjected to a polarizing electric field the general behaviour was much the same in all cases, the main characteristics of the noise response being as follows:

(*a*) Under stresses as low as 1 kV/cm the noise observed was very large and corresponded to thermal-noise input voltages of the order of millivolts.

(*b*) The mean level of the noise was subject to random and violent fluctuations during the few minutes immediately following the application of the polarizing field, and it was impossible to determine a consistent r.m.s. value of the noise during this period.

(*c*) The noise level decayed slowly with time of stress application and was detected in several samples after periods as long as 30 hours.

(*d*) When the noise level at a particular stress had become fairly constant, increase in stress resulted in an increase in noise and a subsequent decay with time. Similar smaller increases followed additional equal increments of stress until at a certain critical stress the noise began to increase much more rapidly. The critical stress varied considerably from sample to sample, and even from day to day on the same sample.

(*e*) Reversal of the field invariably produced a marked initial increase in noise output, the mean level of which fluctuated considerably before again decaying slowly with time.

(*f*) The frequency distribution of the noise, as closely as could be determined, conformed to a $1/f^x$ law where x was approximately 2.

Above 10 kc/s the noise level was immeasurably small.

The specimens investigated exhibited marked differences in noise level under equivalent stresses. In an endeavour to trace the reason for this difference, a systematic comparison was made of the other electrical properties of the samples. It was readily established that the more "noisy" specimens were also inferior electrically in the following respects: (*a*) low insulation resistance, the values encountered falling as low as 10–100 megohms; (*b*) low electric strength, which often resulted in breakdown at stresses as low as 5 kV/cm; (*c*) high loss tangent for this class of material, e.g. 0·1–0·2 at 1 kc/s and 20 volts/cm; and (*d*) very distorted and asymmetric hysteresis loops, resulting presumably from phase distortion produced by the high conductivity.

The most likely cause of the high conductivity was moisture absorbed on the surface and within the bulk of the dielectric, and several specimens were therefore heated to 150° C for several hours and allowed to cool in a dessicator. A considerable improvement in the insulation resistance and noise characteristics resulted, and the hysteresis loops now approached a symmetrical form with much increased maximum polarizations for equivalent stresses. These properties deteriorated again when the specimens were exposed to a moist atmosphere. The general character of the noise also changed, after extended drying, to a more impulsive form, with the magnitude and rate of occurrence of the pulses varying considerably from one specimen to another. It was considered undesirable to undertake a detailed investigation of this residual noise, however, since the evident porosity of the specimens must leave the results open to suspicion.

(8) INVESTIGATIONS ON HIGH-GRADE CERAMICS

(8.1) Description of Specimens

The samples used throughout the work of this Section were specially prepared, particular care being taken to ensure high density (5·4–5·7), complete dryness and freedom from impurity. The capacitors were made in the form of flat plates of 0·8 mm thickness and about 0·5 cm^2 in area. Silver-paste electrodes were fired onto a small central portion of each face, leads were soldered directly to them, and the whole specimen, after drying, was coated with Araldite. The compositions used were pure barium titanate and barium-lead titanate mixtures in the ratios 98 : 2 and 96 : 4. In all cases the insulation resistance was better than 10^{12} ohms at room temperature, even under d.c. stresses as high as 20 kV/cm.

(8.2) General Observations

The noise outputs observed were very low compared with those referred to in Section 7, and corresponded to equivalent thermal-noise input voltages of the order of microvolts. Except where some particular characteristic is more effectively demonstrated by reference to another specimen, which will be clearly indicated, the results stated refer to the barium–lead titanate mixture with a composition in the ratio 98 : 2.

The most readily apparent features of the noise observed, when the specimens were charged slowly to a certain stress, were as follows:

(*a*) It was entirely impulsive in form.

(*b*) The pulses were of two distinct types, differentiated mainly by their respective amplitudes. The smaller pulses occurred mainly below 10kV/cm, and their frequency of occurrence decreased rapidly with time; these will be referred to as the "low stress" noise. The other type of pulse appeared above 14kV/cm; it was much larger in amplitude and seemed to persist as long as the stress was maintained; this will be called the "high stress" noise.

(*c*) The polarity of the pulses as viewed on an oscillograph reversed with the polarity of the applied field.

(*d*) A large proportion of the total noise output at any stress was dependent on the previous electrical history of the specimen.

(*e*) The noise was not affected by immersing the specimens in media such as high-grade transformer oil or silicone fluids, indicating that it resulted from mechanisms internal to the specimen.

(8.3) Investigation of the "Low Stress" Noise

(8.3.1) Noise as a Function of Stress; Hysteresis.

Before the type of experiment to be described was commenced, the specimen was conditioned by taking it through several charge-discharge cycles. The polarizing electric stress was applied for the same period of time in the negative and positive directions. The stress was then increased from zero in steps of 2kV/cm, and the r.m.s. noise outputs, within a bandwidth of 1kc/s around 4·12kc/s, were recorded over equal intervals of time. At the higher stresses where the noise output per unit increase of stress decreased, larger stress increments were employed to ensure measurable outputs. The area under these noise/time records was then used to derive an equivalent noise output per cycle bandwidth at constant gain, per unit increase of stress at each particular value of the applied field. When allowance was made for the slight increase in the basic input thermal-noise level with stress, occasioned by the associated decrease in the effective incremental capacitance of the specimen, the noise output was found to be related to stress in the manner shown in Fig. 11.

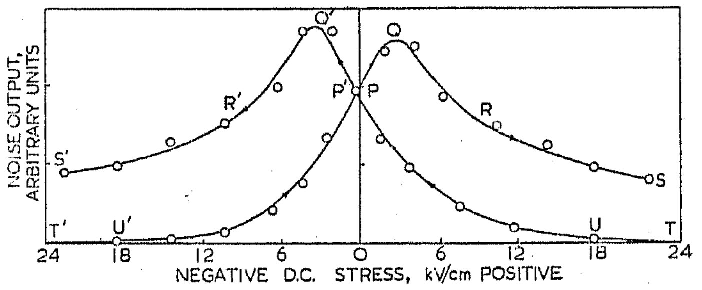


Fig. 11.—Noise output per unit increase of stress plotted against field strength: noise hysteresis.

(8.3.2) Noise Output as a Function of Time.

As with crystals, the noise in the ceramic specimens was by no means confined to the period during which the applied stress was changing but persisted for several seconds after the steady state had been reached. The clearest representation of this noise/time dependence is shown in the records of Fig. 12, where the charging time-constant was 0·1sec. These records also show a marked difference in the duration of the noise decay as between the 98 : 2 barium-lead and pure barium-titanate compositions. The behaviour of the former is shown more fully in the r.m.s. noise/time record of Fig. 13 for a charging time-constant of 6sec.

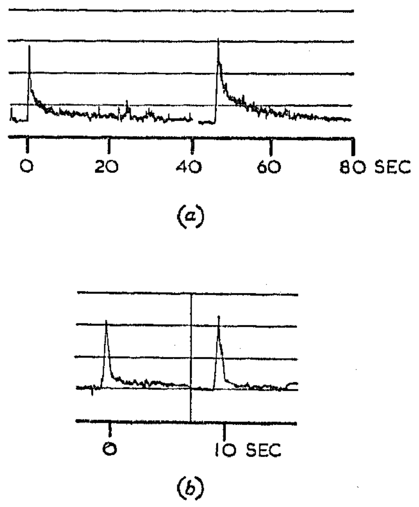


Fig. 12.—Noise records showing the time persistence of the low-stress noise in 98 : 2 $(BaPb)TiO_3$ and $BaTiO_3$ ceramics.

Charging time-constant = 0·1sec.
(*a*) 98 : 2 $(BaPb)TiO_3$.
(*b*) $BaTiO_3$.

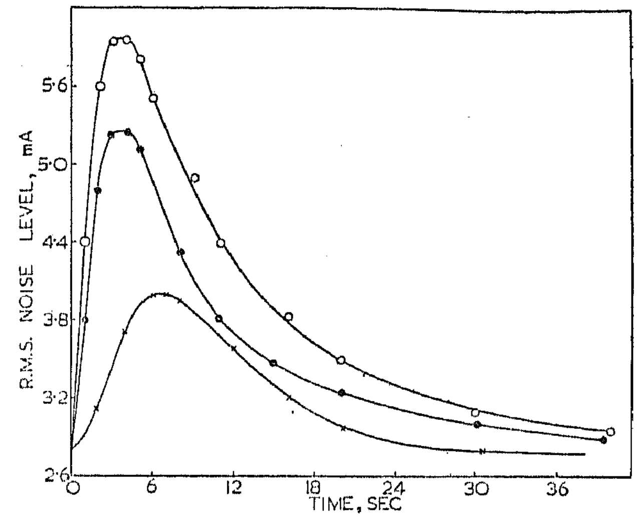


Fig. 13.—Noise/time records for an exponentially increasing stress of time-constant 6sec.

○ ○ ○ 14·0kV/cm.
● ● ● 10·0kV/cm.
× × × 4·5kV/cm.

(8.3.3) History Dependence of the "Low Stress" Noise.

The influence of the previous electrical history of the specimens on the noise response is shown in Fig. 14, where a field of the same polarity was applied a second time to the specimen. This property is even more pronounced than it was in single crystals of barium titanate.

(8.3.4) Noise as a Function of Temperature.

In both the pure barium titanate and the 98 : 2 barium-lead ceramics it was not possible to establish any reliable relationship between noise and temperature. Above 80°C the specimens became very conductive, and large noise outputs persisted as long as the stress was maintained. In form it was essentially impulsive and the number and size of these "conduction" pulses increased rapidly with stress and temperature, thereby masking any other effect which may have been present. Fig. 15 shows the intimate relationship which exists between the conduction current and the r.m.s. noise level, again measured at 4·12kc/s with a 1kc/s bandwidth.

Unlike the results on single crystals, negligible noise was detected if the samples were allowed to cool without an applied field.

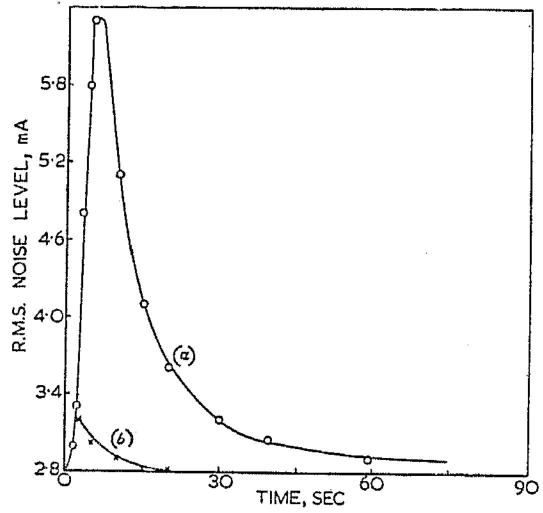


Fig. 14.—History dependence of low stress noise.

98 : 2 (BaPb)TiO_3.
Stress, 14kV/cm.

Time-constant, 6sec.
(*a*) First application of stress.
(*b*) Second application of stress.

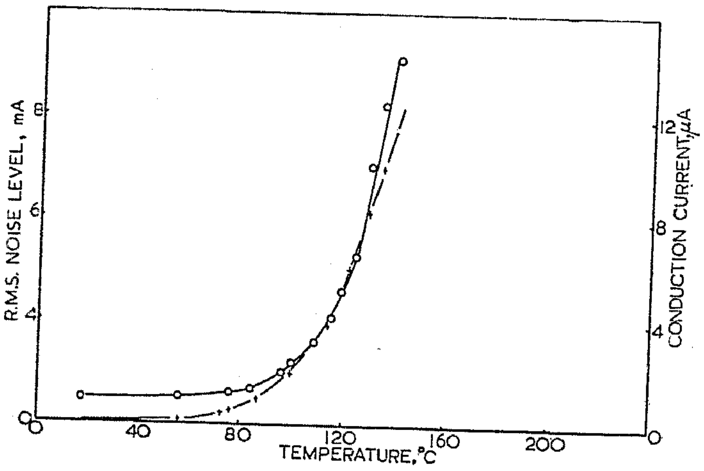


Fig. 15.—Relation between the high-temperature noise and conductivity.

98 : 2 (BaPb)TiO_3.
D.C. stress = 7·5kV/cm.

○○○ R.M.S. noise level.
+++ Conduction current.

(8.3.5) Pulse-Size Distribution and Duration.

The individual pulse amplitudes ranged from 1 to 10 microvolts, while the rise times were of the same order, 10^{-5}sec, as observed with single crystals.

(8.4) Investigation of the Noise appearing above 14kV/cm

At a stress which varied from specimen to specimen but which was generally about 14 kV/cm, a different type of pulse appeared, of much greater amplitude than those occurring at low stresses. They increased in number with the field, showed little sign of diminishing with time, and did not appear to be at all influenced by the previous electrical history of the specimen.

Increasing the temperature of the specimen was found to increase the pulse rate at a given stress, and also to reduce the "inception" stress at which the pulses first appeared. At temperatures where the conductivity had increased sufficiently, it was possible to observe an apparent correlation between these noise pulses, temperature and field strength.

In spite of small time variations of the conductivity it was possible to establish that the logarithm of the conductivity of the 98 : 2 barium-lead composition varied as the inverse of the absolute temperature with an activation energy of the order of 1eV. Its variation with stress may be illustrated by the following figures.

	Temperature			
	70°C		130°C	
Stress, kV/cm	7·5	13·7	7·5	13·7
Log (conductivity), mho/cm	−12·45	−13·40	−11·0	−11·95

Unfortunately it was not possible to obtain reliable corresponding data for the r.m.s. noise because of its violent fluctuations at the higher temperatures.

(9) DISCUSSION OF RESULTS

(9.1) The "Low Stress" Noise

The results obtained at stresses below 10kV/cm may now be examined in the light of the existence of a ferro-electric Barkhausen effect, as demonstrated in the single-crystal investigations.

(9.1.1) Noise as a Function of Stress Hysteresis.

The "butterfly" loop of Fig. 11 is a common one in the relation of various parameters of ferro-electric materials to the applied stress. If the noise observed arises from the Barkhausen effect, the noise/stress relationship of Fig. 11 must be related to the conventional polarization/field-strength characteristic of Fig. 16.

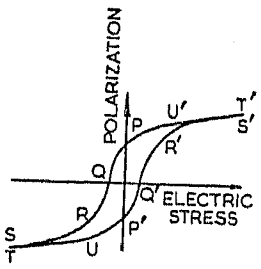


Fig. 16.—Hysteresis loop.

The significant features of this interrelation are:

(*a*) After conditioning in the manner described in Section 8.3.1, zero stress must find the specimen in a state of remanent polarization. In this case, the points P and P′ on the hysteresis loop correspond to P on the noise curve of Fig. 11.

(*b*) The noise output as a function of stress is directly related to dP/dE, reaching a maximum at Q (or Q′) where this has its maximum value, and the field corresponds to the coercive field, E_c.

(*c*) From Q to R the value of dP/dE decreases and the noise output also falls, the latter tending to zero as the polarization approaches saturation.

(*d*) When the field is decreasing from its maximum value the noise output is seen to be related again to the rate of change of polarization. Over the region TU the small decrease in the polarization is mainly due to a decrease in the "induced" or "reversible" polarization, and the noise (which results from "irreversible" changes) is negligible.

(*e*) From U to P′ the increase in dP/dE is due to the irreversible re-randomization of domain moments, and there is a corresponding increase in the noise observed. The pulses constituting this noise are of opposite polarity to those produced when the stress was increasing.

(*f*) From P′ the process is merely repeated producing the mirror image of the section of the curve just considered.

This noise hysteresis could not be satisfactorily established in the measurements on single crystals because of their much more rectangular hysteresis loop and the small stress range over which reversal of the polarization occurs. The more gradual changes inherent in ceramics made the measurement of Fig. 11 possible, and on the above analysis it is seen that a large component of the observed noise can be accounted for in terms of the Barkhausen effect.

(9.1.2) Noise as a Function of Time.

If the noise observed at low stresses is Barkhausen in origin, the pulses which persist after the applied stress has reached its final steady value (Section 8.3.2) must be the result of additional increases or "creep" in the polarization. If this is so, the incremental permittivity in ceramics must also decrease with time under steady d.c. fields.

Typical measurements are shown in Fig. 17. The order of the magnitude of the time effects for the two materials is seen to be much the same as those in the noise record of Fig. 12. The difference between the barium titanate and barium-lead titanate compositions seems to be significant and agrees with the results of Bradfield's piezo-electric constant measurements.

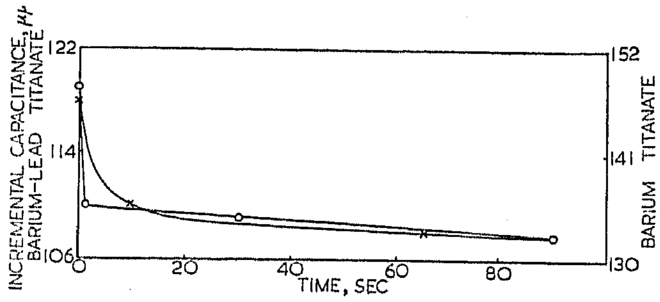


Fig. 17.—Time variation of incremental permittivity.
○○○ Pure barium titanate.
××× 98 : 2 barium-lead-titanate.

(9.1.3) Noise-Maximum/Applied-Stress Relationship.

The series of noise–time curves for exponentially increasing stresses of time-constant 6sec (shown in Fig. 13) may be converted into equivalent noise-stress curves, and according to Section 9.1.1, the noise peak in the latter should then be identifiable with the coercive field E_c. The mean noise level over a small interval of time will depend, however, upon the rate of change of stress over that interval, and for the noise peak to correspond exactly to the point of maximum dP/dE, dE/dt must be constant over the voltage range 0–E_c. With an exponentially increasing stress, ambiguity will only be effectively removed provided that E_{max} is at least equal to $1 \cdot 7E_c$. A similar discrepancy will arise if dE/dt is increased to a value comparable with the response time of the recorder.

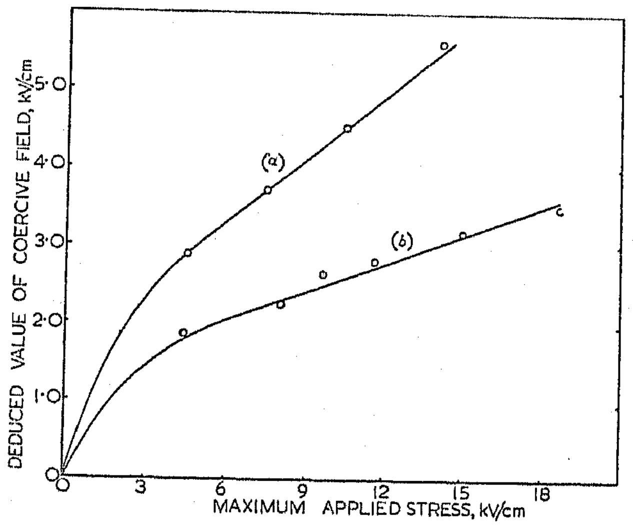


Fig. 18.—Coercive field plotted against applied stress.
(a) From measurements with exponentially increasing field of time-constant 6sec.
(b) From 50c/s hysteresis loops.
98 : 2 (BaPb)TiO$_3$.

If the two sources of error are avoided and the field strength at which the maximum noise output occurs is plotted as a function of the maximum applied stress, curves like (a) in Fig. 18 are obtained. For comparison a curve (b) of E_c as determined from 50c/s hysteresis loops is included. The form of the two curves is obviously the same but the values of E_c given by the 50c/s curve (corresponding to a time-constant of about 10^{-2}sec) are considerably less than those obtained with a changing time-constant of 6sec.

This indication that much larger polarizations are developed in the latter case affords further agreement with Bradfield's results. It must be concluded that the 50c/s hysteresis loop is not a reliable representation of the polarization/field-strength relationship.

(9.2) The "High Stress" Noise

At a stress which varied from specimen to specimen but which was in general around 14kV/cm, the type of noise pulse referred to in Section 2.4 began to appear. The results of Section 8.4.1 suggest that these pulses may be connected with conduction processes in the material and have the same origin as those observed at high temperatures with lower stresses. They may be attributable to electron avalanches or to ionic movements at crystal interfaces; or, notwithstanding that the stress was unidirectional, they may be due to gaseous discharges in small interstices within the material. Much more experimental work will be required to establish which of these mechanisms was the cause of the effects observed or to what extent more than one was operative, and it must suffice at present to mention the results of a few measurements carried out under 50c/s stress using a high-frequency discharge detector.*[14,15]

Discharge pulses first appeared at stresses of 1–2kV/cm and were recognizable as of Barkhausen origin in that the maximum discharge intensity occurred over a stress range corresponding to the steep portion of the hysteresis loop. The individual pulse amplitude was of the order of $10^{-2}\mu\mu$C, and it appeared to increase somewhat with the applied stress—an effect not evident in the d.c. investigation.

At 11–12kV/cm, pulses of amplitude around $1\,\mu\mu$C appeared, and their number and size increased sharply with further increase of stress, until at about 15kV/cm individual pulses had amplitudes as high as $200\,\mu\mu$C. There would seem little doubt that the latter were due to gaseous discharges.

To check that these various effects were not of external origin, the specimens were immersed during test in media such as castor oil and silicone fluid. No change of behaviour was observed, and moreover, a bromide film placed in intimate contact with a specimen under investigation was not marked when developed.

(10) ACKNOWLEDGMENT

The author wishes to express his gratitude to Professor Willis Jackson for his guidance and help during the work described in the paper, to the Bell Telephone Laboratories for supplying the barium titanate crystals mentioned in Section 2 and to Mr. P. Popper of the Mullard Materials Research Laboratory who kindly prepared the samples mentioned in Section 8.1.

(11) BIBLIOGRAPHY

(1) Williams, H. J., and Shockley, W.: *Physical Review*, 1949, 75, p. 178.
(2) Williams, H. J., *et al.*: *ibid.*, p. 155.

---

* These measurements were carried out at the Research Laboratories of the Electrical Research Association by kind permission of Dr. S. Whitehead.

(3) KAY, H. F.: *Acta Crystallographica*, 1948, **1**, p. 229.
(4) MATTHIAS, B., and VON HIPPEL, A.: *Physical Review*, 1948, **73**, p. 1378.
(5) BARKHAUSEN, H.: *Physikalische Zeitschrift*, 1919, **20**, p. 401.
(6) DENNISON, A. T., and WHIDDINGTON, R.: *Proceedings of the Leeds Philosophical and Literary Society*, 1949, **5**, p. 191.
(7) CROSS, L. E., *et al.*: *Nature*, 1949, **163**, p. 635.
(8) NEWTON, R. R., AHEARN, J. A., and MCKAY, K. G.: *Physical Review*, 1949, **75**, p. 103.
(9) TOLSTOY, L. A.: *Zhurnal Tekhnicheskoĭ Fiziki*, 20th August, 1950.
(10) DANIELSON, G. C.: *Acta Crysta*, 1949, **2**, p. 90.
(11) FORSBERGH, P. W.: *Physical Review*, 1949, **76**, p. 1187.
(12) BRADFIELD, G.: National Physical Laboratory Reports, 1950, Vol. 7, Series 9.
(13) BOZORTH, L. M., and DILLINGER, J.: *Physical Review*, 1930, **35**, p. 733.
(14) MOLE, G.: *Proceedings I.E.E.*, Paper No. 1470M, March, 1953 (**100**, Part IIA, p. 276).
(15) MASON, J. H.: *Proceedings I.E.E.*, Paper No. 1471M, March 1953 (**100**, Part IIA, p. 149).

DISCUSSION ON "SPECIAL EFFECTS FOR TELEVISION STUDIO PRODUCTIONS"*

NORTH MIDLAND CENTRE, AT LEEDS, 9TH FEBRUARY, 1954

Dr. G. B. Patchett: Back projection has the disadvantage of requiring a large studio in which to accommodate the projector, together with careful lighting and perspective, but for stationary pictures it may be most satisfactory. Inlay appears to be most attractive, but the authors have been very fair and given the disadvantages of the system as well as its advantages. I imagine that it must be difficult to obtain the same density from the two cameras. Why is a mask used for inlay and an optically-reduced raster for wipes? One difficulty of inlay is the setting-up time during which two cameras are out of action. How long does the setting-up procedure take in a typical case? Overlay appears to be very easy, but it has its difficulties. I assume that with overlay it is not possible to show a shadow on the background of objects in the foreground. One disadvantage of overlay compared with back projection is that the actors are unable to see the background against which they are acting.

It would be most helpful if details of the times when the various new devices, etc., are used by the B.B.C. were published in the technical Press.

Messrs. A. M. Spooner and **T. Worswick** (*in reply*): There is no real difficulty in getting the picture components which comprise the final picture to match in brightness. An essential part of the setting-up procedure consists in the application of a suitable test signal, usually a line sawtooth, to both inputs of the electronic switch. The switch can then be set so that the application of input signals matched in lift and gain will give a matched composite picture.

Setting up an inlay scene takes about one minute, assuming that the procedure has been rehearsed previously. Once the correct-size mask has been prepared and chalk marks have been made for the cameras, it is not difficult to regain the registration obtained on rehearsal.

The reason for using a mask directly on the cathode-ray-tube face for inlay, but a small mask obscuring an optically-reduced raster for wipes, is that, whereas with inlay it is essential to work on a large mask to get the best possible registration, for wipes there is no registration problem and the mechanical arrangements for passing full-size masks across the cathode-ray-tube face would be very cumbersome. Furthermore, in order to get the effect of wipes from different directions, a rotatable-mirror system is used, and this can only turn the reduced raster image.

We should like to confirm that with overlay it is not possible to show a shadow on the background of objects on the foreground. It is true that, with overlay, actors cannot see the background against which they are acting; however, by the use of a picture monitor showing the transmission picture on the studio floor it is possible for them to take up their positions very readily during rehearsal, and in some scenes they may even look at the picture monitor while the scene is being televised.

---

* Paper by A. M. SPOONER and T. WORSWICK (see 1953, 100, Part I, p. 288).

621.396.621:621.396.96

Paper No. 1733 R
Jan. 1955

DESIGN OF A LOGARITHMIC RECEIVER

By S. ROZENSTEIN

(*The paper was first received* 15*th February, and in revised form* 16*th July*, 1954.)

SUMMARY

The paper contains a short survey of the principles of the successive-detection-type logarithmic receiver for pulsed radar, followed by a detailed description of its video problems. A new detection arrangement is described, and a mathematical analysis of its operation under practical conditions is given. A video amplifier including a differentiating circuit for signals of large dynamic range with negligible overshoot is described. Conclusions are drawn for the practical application of the logarithmic receiver with various indicators.

(1) INTRODUCTION

Receiver saturation is one of the limiting effects to the maximum usefulness that may be achieved by a radar system. Any target appearing coincident in time with a mountain, shore, sea or rain clutter or any other large target (usually called a permanent echo), will not be reproduced at the output of an ordinary receiver. The strong signal from the large target will drive the receiver to saturation, and the superimposed wanted target echo will disappear at the video output.

There are several methods for overcoming this difficulty. The simplest solution is the reduction of the receiver gain. When the sensitivity of the receiver is reduced, the strong echo will not saturate the receiver and a smaller echo superimposed on the strong one will be reproduced. The disadvantage of reduced gain is the loss of weak signals. An improved system of gain reduction is the "sensitivity time control" (s.t.c.).[1] The gain of the receiver is reduced to minimum after every transmitted pulse and then gradually raised to the maximum sensitivity at a rate determined by practical considerations particular to the explored area, thereby reducing the saturation possibilities from close permanent echoes. A further improvement is the "instantaneous automatic gain control system" (i.a.g.c.).[2] The sensitivity is normally kept at maximum, and only if the received echo rises above a predetermined level, the sensitivity is automatically reduced according to the strength of the echo. After the strong echo ends, the receiver returns to full sensitivity after a certain recovery time.

A very good method from a practical point of view is to use a logarithmic receiver. The output of this receiver is in a logarithmic relationship to its input, which is achieved by successive detection from each of its amplifying stages. The logarithmic range may be extended as desired and is limited only by the total amplification of the receiver. In the paper a completely new approach to the video problem of the logarithmic receiver is described.

The logarithmic receiver combines the advantages of the above-mentioned systems, absence of saturation even from close permanent echoes as in the s.t.c. system and a shorter recovery time to full sensitivity than in the i.a.g.c. system without its practical disadvantages. These abilities make the logarithmic receiver a very important component in the design of automatic target-following radar systems, moving target indicator (m.t.i.)[3] and other special systems.

(2) THEORETICAL CONSIDERATION

Only a short description will be given here, since the theory of the logarithmic receiver has been fully explained by Croney[4] and Alred and Reiss.[5]

The principle of this receiver consists of detectors (whose individual outputs are added) being connected to the output of every i.f. stage, instead of a single detector following the last i.f. stage as in the linear receiver. This is called the successive-detection system. The block diagram of the linear receiver as compared to the logarithmic one is given in Fig. 1.

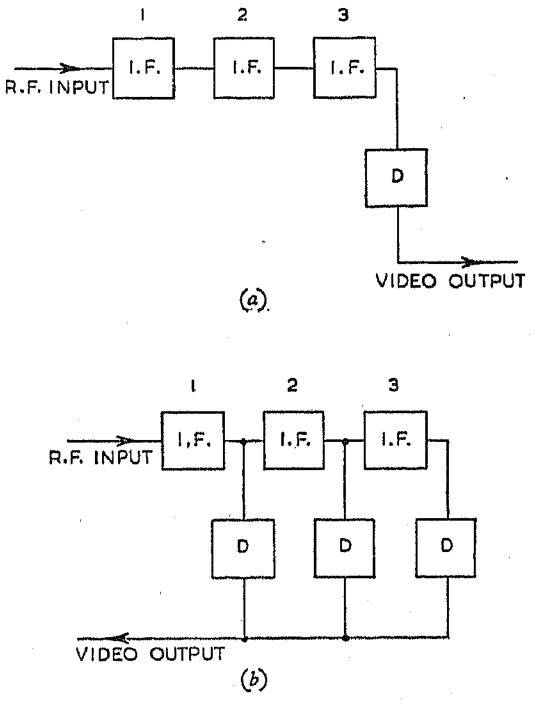


Fig. 1.—Block diagrams.
(*a*) Linear receiver. (*b*) Logarithmic receiver.

Assume the last i.f. stage (3) to be saturated, the detected output voltage to be V_d and the input voltage V_i and assume also that all stages are completely identical and that the output from a saturated tube is constant. Then, in order to saturate the preceding valve (2) the input voltage must be increased A times, where A is the stage amplification, so that input voltage AV_i will increase the total output voltage to $2V_d$. At the saturation level of the preceding valve (1) the input voltage will be A^2V_i and the total output voltage $3V_d$ and so on. We see that while the total input voltage increases exponentially, the total output voltage increases linearly. This results in an approximately logarithmic relationship between output and input. Fig. 2 shows this relationship in a linear and a logarithmic receiver, the input voltage being drawn on a logarithmic scale.

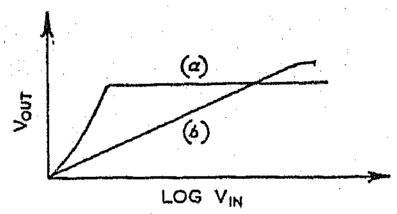


Fig. 2.—Input-output relationship.
(*a*) Linear receiver. (*b*) Logarithmic receiver.

Written contributions on papers published without being read at meetings are invited for consideration with a view to publication.

Mr. Rozenstein is in the Scientific Department, Israeli Ministry of Defence.

(3) PRACTICAL CONSIDERATIONS

It has to be mentioned that all the following discussions are concerned only with pulse receivers and not with continuous-wave receivers.

(3.1) Overload Considerations

A linear receiver reproduces the shape of an r.f. pulse until the input signal is strong enough to saturate the final i.f. stage. A further increase of input signal amplitude will not be reproduced, but the detected output will remain constant. The final i.f. stage should be a valve capable of dealing with as large signals as possible before saturation is reached. If large r.f. pulses drive the final i.f. stage and consequently the previous stages to saturation, only the recovery time, and not the pulse response under overload conditions, is of interest.

In the successive-detection logarithmic receiver, every valve works either as an amplifier or as an output stage, according to the signal strength. From a receiver point of view, the saturation of any stage is not a limiting effect, but its behaviour under overload conditions is of the utmost importance. Constant output above saturation level is one of the basic assumptions that must be fulfilled.

For the design of an i.f. stage under the overload conditions, the circuit and the valve must be carefully chosen. The overload behaviour of valves has already been discussed,[6] but without reference to a suitable circuit.

The standard circuit of a radar i.f. amplifier stage was designed for saving space and decoupling components and for compactness of the stage in order to reduce positive-feedback possibilities. The maximum output voltage is limited by the screen-grid voltage because it also serves as the anode h.t. supply. Therefore it is clear that the circuit is unsuitable as an output stage.

A suitable circuit diagram is shown in Fig. 3. The difference between that circuit and the previous one is the separation of the anode h.t. supply from the screen grid, and its connection to the main h.t. line through a decoupling circuit. While grid current flows, the anode and especially the screen currents rise markedly and the screen by-pass condenser will be discharged, the resultant drop in screen-grid voltage causing a reduction in anode current and therefore a decrease of output. In order to keep the screen-grid voltage constant, as large a screen-grid by-pass condenser as possible—up to space limitations of the practical circuit—must be chosen (with correct layout, the earthed envelope of the by-pass condenser may serve as an electrostatic shielding between adjacent stages). The time-constant of the by-pass condenser with the screen-grid internal impedance during overload must be large compared to the transmitter inter-pulse period. Thus the screen-grid voltage remains practically constant. The time-constant in the cathode circuit must be large compared to the i.f. but small compared to the pulse width, thus ensuring a short recovery time.

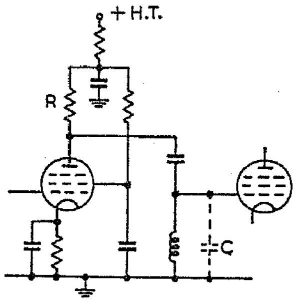


Fig. 3.—I.F. stage for the logarithmic receiver.

Testing of 6AC7 valves in circuit of Fig. 3 failed to give the expected results. This was found to be due to their internal construction. The 6AK5 proved to be satisfactory and was used in the practical receiver. The input-output relationship of 6AC7 compared to 6AK5 connected in the circuit of Fig. 3 is given in Fig. 4.

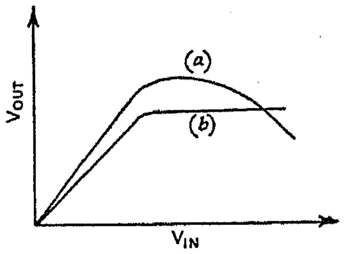


Fig. 4.—Input-output relationships.
(*a*) 6AC7. (*b*) 6AK5.

(3.2) Delay Line

As mentioned in Section 2, the detected outputs from the several i.f. stages must be fed to a common load, where they are added together. The addition must be linear in amplitude or deviation from the logarithmic relationship will result. The leading edges of the video pulses must exactly coincide at the output terminal, otherwise the rise-time of the total output pulse will increase according to the strength of the signal. The circuit diagram of the delay line as suggested by Croney[4] is shown in Fig. 5. The output from the delay line is taken from terminals A.

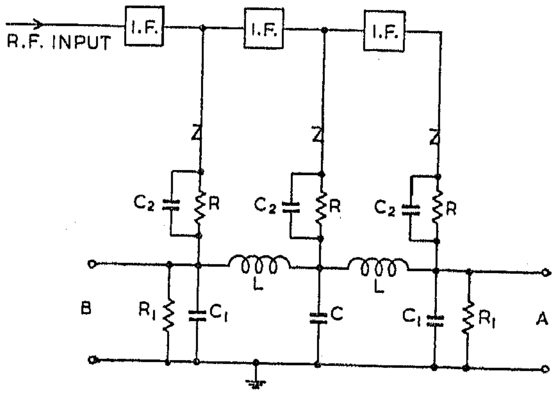


Fig. 5.—Detection arrangement and delay line.

Croney also shows a practical method of calculating the values of *L*, *C*, and C_1, the time-delay per section being equal to the time-delay per stage. Resistance R_1 is equal to the characteristic impedance of the delay line. The line is terminated at both ends to avoid reflections. The common load for all detectors is therefore equal to $\frac{1}{2}R_1$. The ratio $2R/R_1$ is suggested by Croney to be between 5 and 10; the specific ratio depends on the video stage. Since the common load has a shunt capacitance, the voltage divider consisting of decoupling resistor R and the common load resistor, will not function properly at high frequencies unless a capacitor C_2 is added in parallel with R. The time-constants of the two *RC* circuits should be approximately equal. Any section of the delay line which connects the detectors from adjacent stages at the same time forms a filter preventing i.f. feedback through the delay line.

Suppose that the input signal is an ideal square r.f. pulse and is strong enough to saturate the first i.f. stage. The first detector then delivers a video pulse along the delay line. After a finite delay-time the second stage saturates and the video pulse from its detector is also delivered to the line. At that instant the video pulse from the first detector reaches the terminal of the second detector. The amplitudes are added and the leading edges coincide. The resultant pulse continues to travel along the line, repeating the process with the video pulses delivered by the succeeding stages until it reaches the output terminal. The output from the delay line is fed to an *RC* differentiator.

Under practical conditions, graphs drawn from input-output measurements did not show a logarithmic relationship within the expected range, and the video pulse showed an increase in rise time with an increased input pulse-amplitude. It was found that taking the video output from the end of the delay line nearest to the input stage (terminal B) caused an improvement in the rise time of video pulses, resulting in an improved characteristic.

The relationship of input signal voltage to the differentiated output voltage using the output from terminals A compared to terminals B is shown in Fig. 6. The input voltage is drawn on a logarithmic scale.

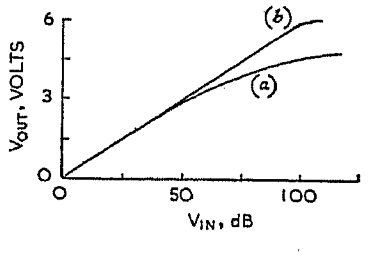


Fig. 6.—Relationship of input signal voltage to differentiated output voltage.
(*a*) Using terminals A (Fig. 5).
(*b*) Using terminals B (Fig 5).

The deviation from the expected results is due to the fact that in practice the input pulse is not an ideal square wave but has a finite rise-time. This may be approximately the same as the total receiver delay-time. In that case, although the final amplitude of the input signal will be strong enough to saturate the first i.f. stage, the rising pulse will be amplified through all stages, increasing in amplitude, while passing through the amplifier at a faster rate than it rises at the input. The final stage will therefore be the first one to be saturated.

A mathematical analysis of this assumption is given in the Appendix. The response of an amplifier of *n* stages to a ramp function is shown to be

$$v_n(t) = \alpha(-1)^n\left(\frac{g_m}{2C}\right)^n\left[\frac{(t/2RC)-n}{(1/2RC)^{n+1}} + \frac{1}{(n-1)!}\,\frac{d^{n-1}}{dp^{n-1}}\left(\frac{v^{pt}}{p^2}\right)_{p=-\frac{1}{2RC}}\right]$$

where α (volts/sec) is the slope of the leading edge and p is the complex frequency. Numerical calculations for an experimental amplifier of three stages were made with $R = 2\,000$ ohms, $C = 11\,\mu\mu\text{F}$, $\alpha = 30 \times 10^6$ (volts/sec) and $g_m = 4\,300$ micromhos. The calculated responses of the three stages were drawn in Fig. 7. The voltage level V' is the input amplitude which is sufficient to saturate the stage. From these curves it is clearly seen that the slope of the leading edge increases from stage to stage so that the third stage reaches saturation before the first.

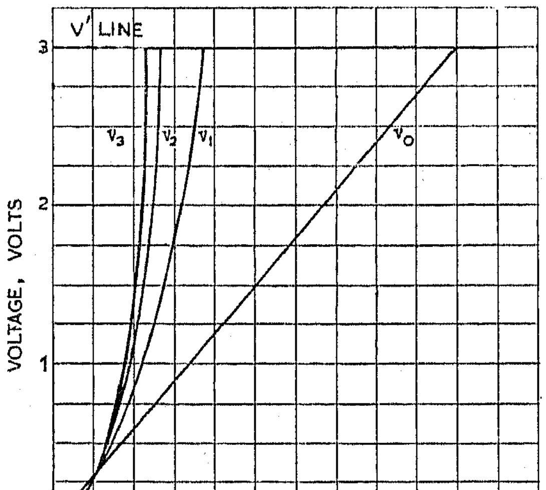


Fig. 7.—Calculated responses of three i.f. stages to a ramp function.

In order to add the leading edges of all the output pulses so that they will coincide, we therefore have to take the output from the end of the delay line which is nearest to the input stage. The delays in the various sections in the delay line are not equal to that in the i.f. amplifier stage (as assumed by Croney[4]), but should be equal to the time intervals $t_1–t_2$, $t_2–t_3$, . . . $t_n–t_{n-1}$ respectively (Fig. 7).

In order to show the difference between this arrangement and the one proposed by Croney, Fig. 8 gives the addition of the outputs from the three stages. Curve A shows the output taken from terminals A (Fig. 5) of a delay line with a time delay of 0·044microsec per section. Curve B shows the output taken from terminals B (Fig. 5), the delay line being calculated according to the method proposed in this paper.

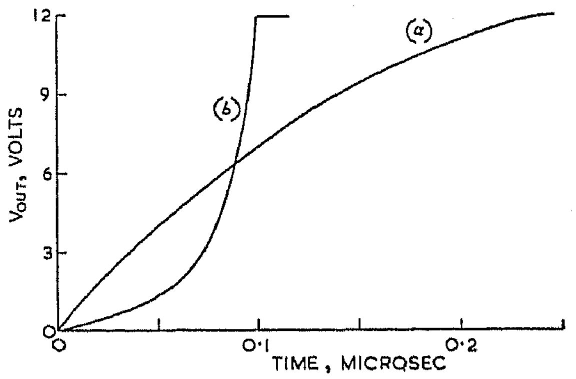


Fig. 8.—Addition of outputs from three stages.
(*a*) From terminals A (Fig. 5).
(*b*) From terminals B (Fig. 5).

(3.3) Differentiator and Video Amplifier

The video output from the line must be amplified by a linear video amplifier capable of dealing with higher input voltages than in the ordinary receiver, for which purpose negative feedback is used in the amplifier. If the maximum input amplitude is higher, the necessary overall video amplification is lower and the amount of feedback may therefore be increased. But as the input circuit to the amplifier is a differentiator, an increase of the maximum input increases the maximum overshoot amplitude (the differentiated trailing edge). The maximum video input is increased by decreasing the ratio of the diode load resistances to the common load. The practical amplification will be the ratio of the maximum pulse amplitude desired for the indicator to the maximum video input pulse. In case the amplification is too high, even when the maximum practical feedback is used (the feedback magnitude limitation is discussed later), the amplification of the amplifier may be controlled by changing the screen voltage of a pentode. The control range is limited at low screen-voltages by the waveform distortion of the pulse and at high screen-voltages by the maximum permitted power dissipation of the screen and the anode. The video amplifier should be able to provide linear amplification up to the highest video pulse, otherwise the logarithmic range of the receiver will be shortened.

As explained in the Introduction, the receiver is most valuable in reproducing a small target superimposed on a large one. The video pulse-width of the large echo is usually much greater than that of the small echo, whose pulse width is equal to the transmitted pulse. That condition is shown in Fig. 9(*a*). It was suggested[5] that a differentiator should be used having a time-constant equal to twice the transmitted pulse-width, as the input

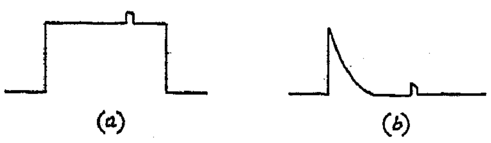


Fig. 9.—Waveform of a small echo superimposed on a large one.
(a) Before differentiation.
(b) After differentiation and clipping.

circuit to the amplifier. The differentiated output (with overshoots eliminated) is shown in Fig. 9(b). The wanted target echo passes through the differentiator without distortion and will be more easily observed at the indicator.

In order to reduce distortion of adjacent echoes, the overshoot time-constant should be as short as possible. In our case the video pulses from the line are of negative polarity; the circuit is therefore designed for elimination of positive overshoots.

The differentiator used in an ordinary receiver is shown in Fig. 10. The ratio between maximum signals and noise at the video input is approximately 5 : 1, and the maximum video pulse-amplitude is approximately 6 volts. A noise amplitude of 1 volt will bias the diode formed by the grid and the cathode of V_1 to cut-off, and its resistance will be large compared to the resistance R. The differentiator time-constant for the leading edge of a negative pulse will be equal to RC, and for the trailing edge to $C[R_dR/(R_d + R)]$, where R_d is the diode resistance during conduction.

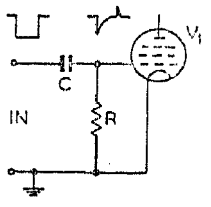


Fig. 10.—Differentiator circuit of a standard receiver.

In the logarithmic receiver the ratio between maximum signals and noise at the video input is approximately 11 : 1 and the maximum video pulse-amplitude is approximately 6 volts. The noise amplitude is therefore lower and will not bias the diode to cut-off, so that its resistance will influence the time-constant of the differentiator even for negative pulses. The resultant parallel resistance will therefore change according to the video input amplitude, so that for small echoes of signal/noise ratio up to 2 : 1 the differentiator time-constant will be lower than for the stronger echoes. That means a reduction in width and amplitude of weak echoes which is contrary to practical requirements. This effect can be avoided by separating the differentiator from the video amplifier by means of a cathode follower, which always presents a high impedance, in parallel with the resistor R. The suggested circuit is shown in Fig. 11. Since the video-amplifier stage is designed for the amplification of negative pulses only, it operates with zero bias. The input to valve 2 is from a low-impedance source; therefore any positive overshoot will not be limited by the grid-cathode conduction effect. Overshoots can be limited by raising the screen voltage as high as possible while still keeping the amplification at the desired value. This holds the valve close to saturation and the anode voltage is very low. Positive input pulses will therefore be limited by saturation of the anode current. The screen voltage must be held constant by a large by-pass condenser.

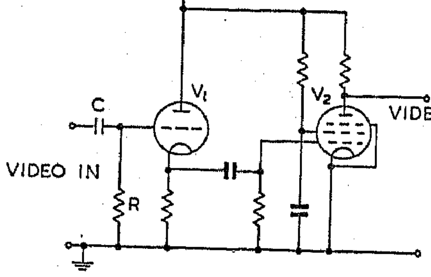


Fig. 11.—Improved differentiator circuit.

The complete circuit diagram of the differentiator and the video amplifier is shown in Fig. 12. Valve 3 is the output stage and valve 4 is the feedback diode. The equivalent circuit of the delay line is represented by the generator V_{DL} and its internal resistance R_{DL}.

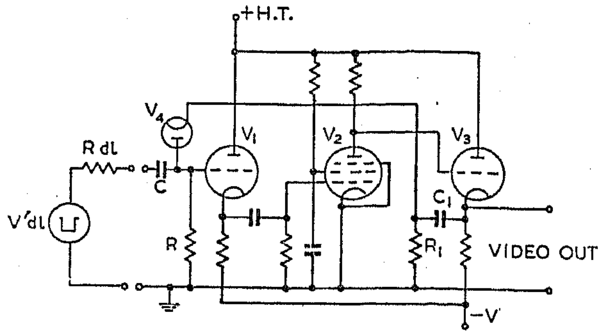


Fig. 12.—Complete circuit diagram of differentiator and video amplifier.

The negative leading edge of the differentiated pulse will be amplified by valve 2 and applied to the output stage, valve 3. A portion of the output is fed back through C_1 and the shunt capacitance of the diode C_d (the diode is cut off) to the input of valve 1. In order to obtain constant feedback action during pulses the time-constant R_1C_1 is made large compared to the transmitter inter-pulse period. The equivalent feedback circuit is shown in Fig. 13. When dealing with the leading edge the impedance Z denotes the diode shunt capacitance C_d. The amount of feedback is determined by the ratio of C_d to C. Owing to the Miller effect, R is shunted by an equivalent capacitance $C_i = C_0(1 + m)$ where m is the overall amplification for negative pulses. A capacitance voltage-divider composed of C and C_i will therefore reduce the effective input to valve 1 and increase the pulse rise-time. The capacitance C is changed in order to obtain the desired feedback ratio. Any change of C must be followed by a change of R to keep RC constant.

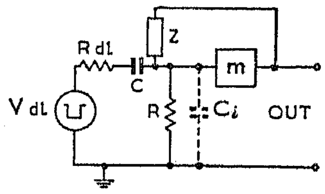


Fig. 13.—Equivalent diagram for feedback circuit.

Since the minimum shunt capacitance of the diode is practically limited, the minimum practical value of C will therefore be $5C_i$. The differentiator time-constant for the leading edge is equal to RC provided that $R \gg R_{dl}$.

The positive trailing edge of the differentiated pulse will drive valve 2 to saturation. All the output voltage from valve 3 is fed back through the diode valve 4 (which is now conducting) to the input of valve 1. In the equivalent circuit (Fig. 13) Z now denotes the forward impedance of the diode. The time-constant of the differentiator will be changed to $C\left(\frac{R_d}{1 + m_1} + R_{dl}\right)$ where m_1 is the overall amplification for positive pulses. This time-constant is very much shorter than that for the leading edge.

The positive overshoot amplitude is practically eliminated at the output and its duration is reduced to a fraction of a microsecond. The recovery time of the logarithmic receiver, after the strongest signals, will therefore depend mainly on the recovery time of the i.f. circuits and not on that of the video amplifier.

Fig. 14 shows the negative video input to the amplifier (shown in Fig. 12) compared to the video output for pulses shorter and longer than the differentiator time-constant.

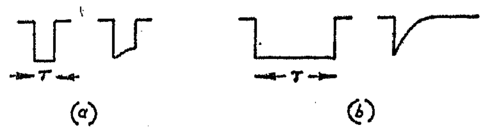


Fig. 14.—Video input and output waveforms of the video amplifier.
(*a*) $RC > \tau$. (*b*) $RC < \tau$.

(4) CONCLUSIONS

It is desirable that the indicator for a logarithmic receiver should have a large dynamic range in order to exploit the ability of the receiver to reproduce both large and small echoes and to permit the observation of the fine structure of large targets and the tracking of a small moving target within the range of a large permanent echo.

Deflection-modulated cathode-ray tubes fit these requirements, while intensity-modulated tubes with their inherent small dynamic range (the ratio of maximum signal/noise amplitude being approximately 3 : 1) are less suitable. But even with these indicators an improvement in discernibility was achieved, giving better discrimination between closely adjacent targets. This is due to the differentiator, which cuts the large target into separate pulses and thereby reduces the spreading caused by large echoes owing to the large illuminated area on the screen.

It was suggested[4, 5] that in order to reduce clutter interference, the logarithmic characteristic should begin well below noise level. But since the extension of the dynamic range of the logarithmic receiver at the same time causes a compression of signal amplitudes on the indicator when using indicators of nearly the same dynamic range as for a linear receiver, a reduction in the amplitude of small echoes causes a reduction of signal/noise ratio on the indicator and in practice reduces the detectability of distant targets. It is therefore necessary to provide an arrangement ("lin-log" switch[4]) for switching the receiver to linear operation when searching for distant targets and to logarithmic operation when the tracked target enters an area of permanent echoes or clutter.

A practical logarithmic receiver was built in a plug-in form and inserted in the receiver unit of a radar system. Excellent results as regards clutter elimination and tracking of targets through clutter areas were obtained.

(5) ACKNOWLEDGMENTS

The author wishes to express his thanks to Dr. E. D. Bergman, Director of the Scientific Department, Israeli Ministry of Defence, for permission to publish the paper. He wishes to thank Mr. E. Gross for his assistance in the practical laboratory work and his encouragement and guidance in preparing the paper. The author is grateful to Messrs. A. F. Fischman, M. Zakhaim and J. Shekel for many useful discussions and fruitful suggestions while the paper was being prepared.

(6) REFERENCES

(1) van Voorhis, S. N.: "Microwave Receivers" (McGraw-Hill Book Co., Inc., New York, 1948, Radiation Laboratories Series, Volume 23, p. 251).
(2) *Ibid.*, p. 248.
(3) *Ibid.*, p. 583.
(4) Croney, J.: "A Simple Logarithmic Receiver," *Proceedings of the Institute of Radio Engineers*, 1951, **39**, p. 807.
(5) Alred, R. V., and Reiss, A.: "An Anti-Clutter Radar Receiver," *Journal I.E.E.*, 1948, **95**, Part III, p. 459.
(6) van Voorhis, S. N.: *Op. cit.*, p. 184.
(7) Valley, Wallman: "Vacuum Tube Amplifiers" (McGraw-Hill Book Co., Inc., New York, N.Y., 1948, Radiation Laboratories Series, Volume 18, p. 64).

(7) APPENDIX

(7.1) Analysis of the Response of a Saturated I.F. Amplifier of *n* Stages

The response of an amplifier of n stages to a unit step-function is well known.[7] In practice, however, the r.f. pulse fed into an amplifier has a finite rise-time. In contrast with a linear amplifier, where the rise time is usually computed, it is the slope of the leading edge which is of main concern here, because the top of the pulse is limited by saturation.

Fig. 15(*a*) is the equivalent circuit of a practical i.f. stage as shown in Fig. 3. The bandwidth of this circuit is approximately 7 Mc/s and the carrier frequency 30 Mc/s. The low bandwidth/carrier-frequency ratio enables us to simplify the analysis by computing the response of the equivalent low-frequency circuit [Fig. 15(*b*)] to the envelope of the r.f. pulse.

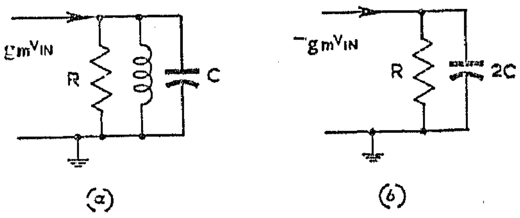


Fig. 15.—Equivalent circuits.
(*a*) Of Fig. 3.
(*b*) Low-frequency equivalent.

In order to simplify the analysis we will assume that the input voltage $v_0(t)$ is a ramp function of a slope α (volts/sec). The time interval t_1 is the rise-time of the pulse (Fig. 7). The voltage level V' is the input amplitude sufficient to saturate the stage. We shall deal only with the leading edge since that is the only part of the pulse needed for the ranging of targets in the radar system.

The Laplace transform of the input voltage $v_{in}(t)$ is given by

$$\mathscr{L}v_0(t) = \mathscr{L}\alpha t = \alpha\frac{1}{p^2} \quad . \quad . \quad . \quad . \quad . \quad (1)$$

and the impedance of the circuit in Fig. 15(*b*) is given by

$$Z(p) = \frac{1}{2C}\frac{1}{[p + 1/(2RC)]} \quad . \quad . \quad . \quad . \quad (2)$$

The output voltage for one stage will be

$$v_1(t) = -\,v_0 g_m Z \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

and for n stages

$$v_n(t) = (-1)^n v_0(t)(g_m Z)^n \quad . \quad . \quad . \quad . \quad (4)$$

Expressed in the Laplace transform it becomes

$$\mathscr{L}v_n(t) = (-1)^n\alpha\left(\frac{g_m}{2C}\right)^n\frac{1}{p^2[p+1/(2RC)]^n}$$
$$= (-\,1)^n\alpha\beta^n\frac{1}{p^2(p+\gamma)^n} \quad . \quad . \quad . \quad . \quad (5)$$

where

$$\left.\begin{aligned}\beta &= \frac{g_m}{2C}\\ \gamma &= \frac{1}{2RC}\end{aligned}\right\} \quad . \; . \; . \; . \; . \; . \; . \quad (6)$$

The response of the output voltage after n stages will then be solved by taking the inverse transform of eqn. (5):

$$v_n(t) = \mathscr{L}^{-1}v_n(p) = (-1)^n\alpha\beta^n\mathscr{L}^{-1}\frac{1}{p^2(p+\gamma)^n} \quad . \quad (7)$$

The inverse transformation will be evaluated as

$$\mathscr{L}^{-1}f(p) = \frac{1}{2\pi j}\int_{c-j\infty}^{c+j\infty} f(p)v^{pt}dp \quad . \; . \; . \; . \quad (8)$$

Using Jordan's theorem,

$$\int_{-\infty}^{+\infty} f(p)v^{pt}dp = 2\pi j\,\Sigma\,R_{es}^{+} \quad . \; . \; . \; . \quad (9)$$

where ΣR_{es}^{+} denotes the sum of the residues of $f(p)v^{pt}$ at its poles in the left half-plane. From eqn. (7) we see that the poles are $p = 0$ and $p = -\gamma$. We can therefore rewrite eqn. (8)

$$v_n(t) = (-1)^n\alpha\beta^n\frac{1}{2\pi j}\left[\oint\frac{v^{pt}/(p+\gamma)^n}{(p-0)^2}dp + \oint\frac{v^{pt}/p^2}{(p+\gamma)^n}dp\right] \quad . \; . \; . \; . \quad (10)$$

From the formula of the derivative of an analytic function we get

$$\oint\frac{f(w)}{(w-Z)^{n+1}}dw = \frac{2\pi j}{n!}f^n(Z)_{w=z} \quad . \; . \; . \quad (11)$$

The eqn. (10) can therefore be developed by eqn. (12) to

$$v_n(t) = (-1)^n\alpha\beta^n\left\{\frac{d}{dp}\left[\frac{v^{pt}}{(p+\gamma)^n}\right]_{p=0} + \frac{1}{(n-1)!}\,\frac{d^{n-1}}{dp^{n-1}}\left(\frac{v^{pt}}{p^2}\right)_{p=-\gamma}\right\} \quad . \; . \; . \; . \quad (12)$$

Introducing the terms β and γ from eqn. (6) yields the general formula for the output voltage of the nth stage:

$$v_n(t) = (-1)^n\alpha\left(\frac{g_m}{2C}\right)^n\left\{\frac{(t/2RC) - n}{[1/(2RC)]^{n+1}} + \frac{1}{(n-1)!}\,\frac{d^{n-1}}{dp^{n-1}}\left(\frac{v^{pt}}{p^2}\right)_{p=-\frac{1}{2RC}}\right\}. \quad (13)$$

It must be noted that the output voltage of any stage is limited at the level V' (Fig. 7) and remains constant thereafter.

DISCUSSION ON "A METHOD OF DESIGNING TRANSISTOR TRIGGER CIRCUITS"[*]

Mr. H. M. S. Smith (*communicated*): Reference is made to the problem of "surplus holes" in association with the minimum decay time of pulses. The authors' remarks have all been concerned with the point-contact type of transistor, and it would be interesting to know whether they are equally applicable to junction-type transistors. Have the latter types a frequency response similar to the contact types?

Considering the Hall effect in germanium, one would imagine that transistors would have their characteristics modified by a magnetic or electric field. This matter is of interest in the use of transistors near heavy-current-carrying conductors or in electric fields of large potential gradient.

Prof. F. C. Williams and **Mr. G. B. B. Chaplin** (*in reply*): The frequency response of presently-available junction transistors in the earthed-base connection is inferior to that of the point-contact transistor under similar conditions, by a factor of 3 or more. The response to an input current pulse sufficiently large to produce "bottoming" is also inferior in that both the rise and fall times of the collector are increased by a similar factor.

It follows from the pulse response that the conventional cross-coupled junction-transistor two-state circuit has relatively long transition times, since the time required to change its state includes not only the "switch on" time of one transistor but also the "switch off" time of the other. When this is compared with the "switch on" time of the point-contact single-transistor two-state circuit, it is slower by a factor of about 10. However, junction transistors with a substantially higher frequency response are promised by the manufacturers in the near future.

We have had little experience of the effect of magnetic and electric fields on transistor characteristics, but the shielding provided by the manufacturers appears to be adequate under normal conditions.

---

[*] Paper by Prof. F. C. Williams and G. B. B. Chaplin (see 1953, 100, Part III, p. 228).

621.372.8

Paper No. 1763 R
Jan. 1955

PROPAGATION IN CURVED AND TWISTED WAVEGUIDES OF RECTANGULAR CROSS-SECTION

By L. LEWIN, Associate Member.

(*The paper was first received* 8*th June, and in revised form* 1*st September*, 1954.)

SUMMARY

The propagation in curved and twisted rectangular waveguides is investigated by putting the wave equation in a form in which the co-ordinates in a waveguide cross-section are also the independent variables in the differential equation. The wavelength is found for the E-curved and H-curved guides, and for twisted guides. In the latter case there is a degeneracy when the guide sides become equal. Curves are presented for the correction to the wavelength in a straight guide caused by the bending or twisting.

LIST OF SYMBOLS

a, b = Guide cross-section dimensions (a normally greater than b).
e = Correction term [see eqn. (46)].
h, h_0, h_1, etc. = Magnetic field and expansion terms.
i_x, i_X, etc. = Unit vectors.
k, k' = $2\pi/\lambda$ and $2\pi/\lambda_g$, respectively.
m, n = Positive integers.
$p = 2\pi/L$.
r = Polar co-ordinate.
s = Axial co-ordinate.
xyz = Rectangular co-ordinates.
y = Guide cross-section co-ordinate.
$ABCD$, $A'B'C'D'$, $A_0A_1A_2A_{mn}$ = Various constants and coefficients.
E, E_0, E_1, E_X, etc. = Electric-field components and expansion coefficients.
F, G = Functions of ϕ.
H, H_x = Magnetic field and components.
L = Length of twisted guide for 360° twist.
$M = [1 + p^2(x^2 + y^2)]^{\frac{1}{2}}$.
R = Radius of curvature of guide mid-section.
S, S_0 = Series.
X, Y, Z = Helical co-ordinates.
α, β = Expansion coefficients of the propagation coefficient.
γ = Propagation coefficient in twisted or curved guide = 2π/wavelength.
ε = Base of Naperian logarithms.
$\epsilon_n = 1$ if $n = 0$, otherwise 0.
λ, λ_g = Free-space and straight-guide wavelength, respectively.
$\lambda_E, \lambda_H, \lambda_T$ = Wavelength in E-curved, H-curved, and twisted guides respectively.
η = Wave impedance of free space.
θ = Polar co-ordinate.
ψ, ψ_0, ψ_1, etc. = Solution of wave equation and expansion coefficients.
$\phi = \frac{1}{2}\pi b/a$.

(1) INTRODUCTION

Recent years have seen an increased use, especially at the shortest wavelengths, of curved and twisted lengths of waveguide for joining off-set components. When such a guide is also used to provide a delay path it becomes important to know the change of phase introduced by the bending or twisting. This aspect is investigated theoretically in the following Sections.

(2) METHOD OF APPROACH

Two of the configurations—the circular E-bend and the circular H-bend—can be treated exactly, using Bessel functions of general order. However, the process of finding the propagation coefficient in the guide involves the solution of a transcendental equation, which can be done only by a rather awkward manipulation of their asymptotic expansions for large order and argument. An alternative method, used by Jouguet[1] in an examination of mode degeneration in bent circular guides, involves the setting-up of the differential equations for the field components in a form which permits an approximate solution in a form similar to that for the straight guide. Introducing as a small quantity the expression $1/R$, where R is the radius of curvature of the mid-section of the curved guide, an expansion in powers of $1/R$ is obtained. Since the propagation in the guide cannot depend on the direction of bending, the solution for the propagation coefficient must be an even function of R, and hence terms to an order of at least $1/R^2$ have to be retained in the solution.

For the twisted guide a similar procedure is adopted, the role of R being replaced by the length L in which the guide cross-section has rotated a full 360°. An expansion at least up to terms involving $1/L^2$ is required, since the direction of twist, determined by the sign of L, cannot affect the value of the propagation coefficient. For the twisted guide it is necessary to set up the wave equation in helical co-ordinates.

(3) THE CIRCULAR E-PLANE BEND

It is well known[2] that problems involving a guide of finite breadth, in which no variation in field across the breadth occurs except for the usual sinusoidal variation of the main guide mode, can be treated by considering first a guide of infinite breadth—a parallel-plate line—and then at the end of the analysis replacing the free-space wavelength λ, where it occurs in the formula, by λ_g, the wavelength in the finite-breadth guide. This simplifying procedure will be adopted here.

Fig. 1 shows a guide of width b bent into an arc of a circle of radius of curvature R at the mid-section. The cross-section is determined by the variation of the co-ordinate y from $+\frac{1}{2}b$ to $-\frac{1}{2}b$, and the axial co-ordinate at the mid-section is s. In terms of polar co-ordinates (r, θ) we have

$$s = R\theta \text{ and } r = R + y \quad . \quad . \quad . \quad . \quad . \quad (1)$$

The wave equation can be set up for the field components E_r, E_0

Written contributions on papers published without being read at meetings are invited for consideration with a view to publication.
Mr. Lewin is with Standard Telecommunication Laboratories, Ltd.

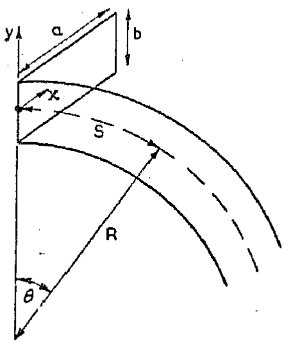


Fig. 1.—Circular E-bend guide and co-ordinate system.

and H_x. Denoting the latter simply by H, its differential equation is readily found to be

$$r^2\frac{\partial^2 H}{\partial r^2} + r\frac{\partial H}{\partial r} + r^2k^2H + \frac{\partial^2 H}{\partial \theta^2} = 0 \quad . \quad . \quad . \quad (2)$$

where $k = 2\pi/\lambda$.

In terms of y and s we get, after dividing through by R^2,

$$(1 + y/R)^2\frac{\partial^2 H}{\partial y^2} + (1 + y/R)\frac{1}{R}\frac{\partial H}{\partial y} + (1 + y/R)^2k^2H + \frac{\partial^2 H}{\partial s^2} = 0 \quad . \quad . \quad . \quad . \quad (3)$$

We seek a solution varying as $\varepsilon^{-j\gamma s}$, where γ is to be determined. The boundary conditions on H are that $E_\theta \propto \partial H/\partial y$ should vanish at $y = \pm\frac{1}{2}b$. Setting $H = h\varepsilon^{-j\gamma s}$, where h is independent of s,

$$(1 + y/R)^2\frac{d^2h}{dy^2} + (1 + y/R)\frac{1}{R}\frac{dh}{dy} + (1 + y/R)^2k^2h - \gamma^2h = 0 \quad . \quad . \quad . \quad . \quad (4)$$

For infinite R the equation becomes simply $d^2h/dy^2 + (k^2 - \gamma^2)h = 0$. If the guide will support only the dominant mode, as will be assumed, the only solution which satisfies $dh/dy = 0$ at $y = \pm\frac{1}{2}b$ is $h =$ constant. Hence $\gamma = k$, the expected solution for a straight guide.

A solution of eqn. (4) in inverse powers of R would take the form

$$\left.\begin{aligned} h &= A(h_0 + h_1/R + h_2/R^2 + \ldots) \\ \gamma^2 &= k^2(1 + \alpha/R + \beta/R^2 + \ldots) \end{aligned}\right\} \quad . \quad . \quad . \quad (5)$$

Here, A is any constant; but it would be confusing to consider a form for A which itself involved inverse powers of R. Such a form, on carrying out the indicated multiplication would, for example, introduce an arbitrary term involving h_0 in the expression for h_1. In order to deal with a unique solution, it will be assumed that h_1, h_2, etc., will not involve h_0 explicitly. h_0, of course, is the solution for a straight guide and can be taken as unity. On substituting eqn. (5) into eqn. (4), and equating powers of $1/R$, the following equation is found for h_1:

$$d^2h_1/dy^2 + 2yk^2 - k^2\alpha = 0 \quad . \quad . \quad . \quad . \quad (6)$$

Integrating once it is seen that

$$dh_1/dy = \text{constant} - k^2(y^2 - \alpha y) \quad . \quad . \quad . \quad (7)$$

The constant and α are to be determined by $dh_1/dy = 0$ at $y = \pm\frac{1}{2}b$. This gives $dh_1/dy = k^2(b^2/4 - y^2)$ and $\alpha = 0$, the latter as anticipated in Section 2.

A further integration yields

$$h_1 = k^2y(b^2/4 - y^2/3) \quad . \quad . \quad . \quad . \quad . \quad (8)$$

with no further integration constant, since this would have the effect, as discussed above, of including an h_0 term in h_1. The equation for h_1, found by equating coefficients of $1/R^2$ in eqns. (4) and (5), is

$$d^2h_2/dy^2 - 4k^2y^2 + k^2b^2/4 + 2y^2k^4(b^2/4 - y^2/3) - k^2\beta = 0 \quad . \quad (9)$$

An integration gives

$$dh_2/dy = C + k^2b^2y/4 - 4k^2y^3/3 + k^4b^2y^3/6 - 2k^4y^5/15 - k^2\beta y \quad . \quad . \quad . \quad (10)$$

The vanishing of dh_2/dy at $y = \pm\frac{1}{2}b$ gives $C = 0$ and

$$\beta = -b^2/12 + k^2b^4/30 \quad . \quad . \quad . \quad . \quad (11)$$

Calling λ_E the wavelength in the E-plane curved guide, it is readily seen, from eqns. (11) and (5) with $\gamma = 2\pi/\lambda_E$, that[3]

$$\frac{1}{\lambda_E^2} = \frac{1}{\lambda_g^2}\left\{1 - \frac{b^2}{12R^2}\left[1 - \frac{2}{5}\left(\frac{2\pi b}{\lambda_g}\right)^2\right]\right\} \quad . \quad . \quad (12)$$

which is correct to terms involving $1/R^2$. In obtaining this equation, λ has been replaced by λ_g, so that the result now refers to a guide of finite breadth.

A continuation of the process will yield higher (even) powers of $1/R$. However, even for so small a radius of curvature as $R = b$, the correction to the straight-guide wavelength shown by eqn. (12) is small—about 3% or less for the usual guide sizes and wavelength ranges. The value $\lambda_g = 2\pi b\sqrt{(2/5)} = 3{\cdot}97b$, for which the correction term vanishes, is within the usual range. The quantity $[1 - 2/5(2\pi b/\lambda_g)^2]$ is plotted as a function of b/λ_g in Fig. 2.

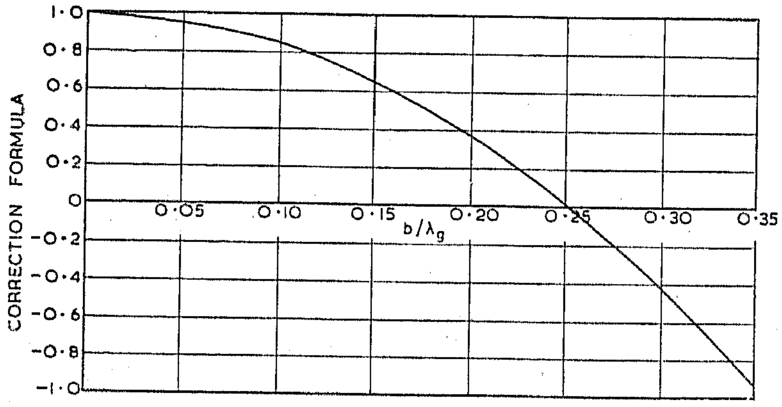


Fig. 2.—Plot of correction function $f(b/\lambda_g) = 1 - \frac{2}{5}\left(\frac{2\pi b}{\lambda_g}\right)^2$ against b/λ_g for E-plane circular bend.

(4) THE CIRCULAR H-PLANE BEND

Fig. 3 shows a guide of breadth a bent into an arc of a circle of radius of curvature R at the mid-section. The cross-section is determined by the variation of the co-ordinate y from $+\frac{1}{2}a$ to

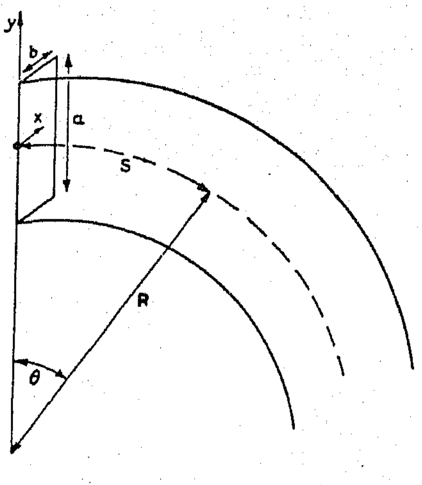


Fig. 3.—Circular H-bend guide and co-ordinate system.

$-\frac{1}{2}a$, and the axial co-ordinate at the mid-section is s. As before, y and s are expressible in terms of polar co-ordinates r and θ by eqn. (1). The field components are H_r, H_θ and E_x. The latter is readily shown to satisfy the differential equation

$$r^2\frac{\partial^2 E_x}{\partial r^2} + r\frac{\partial E_x}{\partial r} + r^2k^2E_x + \frac{\partial^2 E_x}{\partial \theta^2} = 0 \quad . \quad . \quad (13)$$

We seek a solution of the form $E_x = E\varepsilon^{-j\gamma s}$, where E is independent of s. Analogously to eqn. (4) we find the equation for E as

$$(1 + y/R)^2\frac{d^2E}{dy^2} + (1 + y/R)\frac{1}{R}\frac{dE}{dy} + (1 + y/R)^2k^2E - \gamma^2E = 0 \quad . \quad . \quad . \quad . \quad (14)$$

The boundary conditions on E are such that it should vanish at $y = \pm\frac{1}{2}a$. For an infinite value of R the equation becomes simply $d^2E/dy^2 + (k^2 - \gamma^2)E = 0$. If the guide will support only the dominant mode, as will be assumed, the only solution of this which satisfies $E = 0$ at $y = \pm\frac{1}{2}a$ is $E = \cos(\pi y/a)$; whence $\gamma^2 = k^2 - \pi^2/a^2$, the expected solution for the straight guide.

A solution of eqn. (14) in inverse powers of R would take the form

$$\left.\begin{aligned} E &= A_0(E_0 + E_1/R + E_2/R^2 + \ldots) \\ \gamma^2 &= k'^2(1 + \alpha/R + \beta/R^2 + \ldots) \end{aligned}\right\} \quad . \quad . \quad (15)$$

Here $k'^2 = k^2 - \pi^2/a^2 = (2\pi/\lambda_g)^2$, where λ_g is the wavelength in the straight guide and $E_0 = \cos(\pi y/a)$. As before, in order to deal with a unique solution, E_1, E_2, etc., will be taken not to involve E_0 explicitly. On substituting eqn. (15) into eqn. (14) and equating powers of $1/R$, equations are found for the E's.

For E_1 we get

$$\frac{d^2E_1}{dy^2} + \frac{\pi^2}{a^2}E_1 + 2y\frac{d^2E_0}{dy^2} + \frac{dE_0}{dy} + 2yk^2E_0 - k'^2\alpha E_0 = 0 \quad . \quad (16)$$

The solution is

$$E_1 = \frac{a^2}{2\pi^2}\left[- k^2yE_0 + k'^2\left(y^2 - \frac{a^2}{4}\right)dE_0/dy\right]$$
$$- \frac{1}{2}\frac{a^2}{\pi^2}k'^2\alpha y\frac{dE_0}{dy} + A\cos(\pi y/a) + B\sin(\pi y/a) \quad . \quad (17)$$

Since E_1 must not involve $E_0 = \cos(\pi y/a)$ explicitly, $A = 0$. The vanishing of E_1 at $y = \pm\frac{1}{2}a$ gives $B = 0$ and $\alpha = 0$, the latter showing the expected absence of terms in $1/R$ from the propagation coefficient.

The equation for E_2 becomes

$$\frac{d^2E_2}{dy^2} + \frac{\pi^2}{a^2}E_2 + y^2\frac{d^2E_0}{dy^2} + 2y\frac{d^2E_1}{dy^2} + y\frac{dE_0}{dy} + \frac{dE_1}{dy}$$
$$+ y^2k^2E_0 + 2yk^2E_1 - k'^2\beta E_0 = 0 \quad . \quad (18)$$

This can be put in the form

$$\frac{d^2E_2}{dy^2} + \frac{\pi^2}{a^2}E_2 = (A + By^2)\cos(\pi y/a) + y(C + Dy^2)\sin(\pi y/a) \quad . \quad . \quad . \quad . \quad (19)$$

on substituting for E_1 from eqn. (17), since $E_0 = \cos(\pi y/a)$.

The solution, apart from the complementary function, is of the form

$$E_2 = y^2(A' + B'y^2)\cos(\pi y/a) + y(C' + D'y^2)\sin(\pi y/a) \quad . \quad (20)$$

where, by substitution into eqn. (19), it is found that

$$\left.\begin{aligned} A' &= \frac{a}{4\pi}(- C + Ba/\pi + D12a^2/8\pi^2) \\ B' &= - Da/8\pi \\ C' &= \frac{a}{2\pi}(A + Ca/2\pi - Ba^2/2\pi^2 - D3a^3/4\pi^2) \\ D' &= Ba/6\pi + Da^2/4\pi^2 \end{aligned}\right\} \quad . \quad (21)$$

The boundary conditions, $E_2 = 0$ at $y = \pm\frac{1}{2}a$, plus the uniqueness requirement calling for E_0 not to appear explicitly in the solution leads to the disappearance of the complementary function, and to the equation.

$$C' + D'a^2/4 = 0 \quad . \quad . \quad . \quad . \quad . \quad (22)$$

C' and D' can be put in terms of A, B, C and D from eqn. (21), and the latter are found on comparing eqns. (18) and (19), using eqn. (17) to express E_1 in terms of $E_0 = \cos(\pi y/a)$. After some simplification, the following equation is obtained for β:

$$12k'^2\beta = k'^4a^4(15 - \pi^2)/\pi^4 + 3k^2a^2/\pi^2 + k'^2a^2(15 - 3\pi^2)/\pi^2 \quad . \quad . \quad . \quad . \quad . \quad (23)$$

Calling λ_H the wavelength in the H-plane curved guide, it is found from eqns. (23) and (15), with $\gamma = 2\pi/\lambda_H$, that[4]

$$\frac{1}{\lambda_H^2} = \frac{1}{\lambda_g^2} + \frac{1}{24R^2}\left[1 - \frac{12 + \pi^2}{2\pi^2}\left(\frac{2a}{\lambda}\right)^2 + \frac{15 - \pi^2}{2\pi^2}\left(\frac{2a}{\lambda}\right)^4\right] \quad . \quad (24)$$

which is correct down to terms involving $1/R^2$.

A continuation of the process would yield higher (even) powers of $1/R$. However, even for so small a radius of curvature as $R = a$, the correction to the straight-guide wavelength shown by eqn. (24) is small—about 3% or less for the usual guide sizes and wavelength ranges. The values $\lambda = 1{\cdot}161a$ and $1{\cdot}755a$, which lie near the extremes of the usual range, make the correction term in $1/R^2$ vanish. The correction term takes its largest magnitude of $-0{\cdot}0076/R^2$ at $\lambda = 1{\cdot}37a$. The function $\left[1 - \frac{12 + \pi^2}{2\pi^2}\left(\frac{2a}{\lambda}\right)^2 + \frac{15 - \pi^2}{2\pi^2}\left(\frac{2a}{\lambda}\right)^4\right]$ versus $2a/\lambda$ is plotted in Fig. 4.

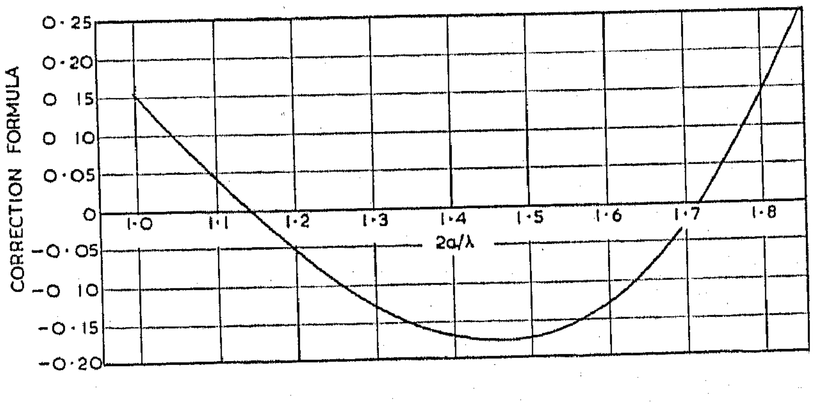


Fig. 4.—Plot of correction function

$$g(a/\lambda) = \left[1 - \frac{12 + \pi^2}{2\pi^2}\left(\frac{2a}{\lambda}\right)^2 + \frac{15 - \pi^2}{2\pi^2}\left(\frac{2a}{\lambda}\right)^4\right]$$

against $2a/\lambda$ for H-plane circular bend.

(5) THE TWISTED GUIDE

Fig. 5 shows a co-ordinate system in which mutually orthogonal axes XY are rotated with respect to fixed axes xy, the angle of rotation being proportional to the axial co-ordinate z. Write

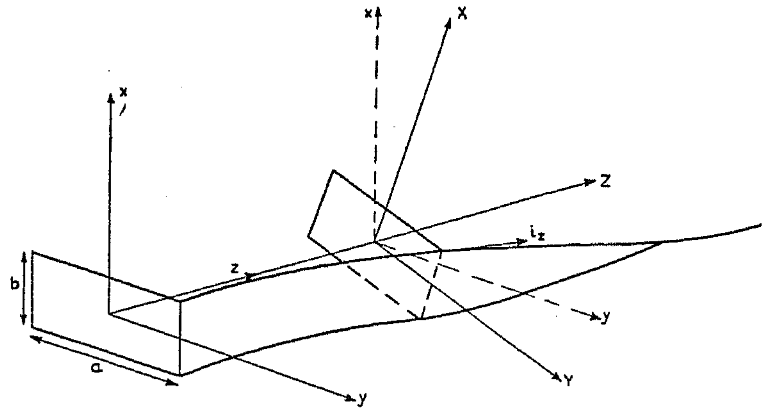


Fig. 5.—Twisted guide and helical co-ordinate system.

p for $2\pi/L$, where L is the length in which the XY axes rotate 360°. Then we have the transformation

$$\left.\begin{aligned} X &= x\cos(pz) + y\sin(pz) \\ Y &= y\cos(pz) - x\sin(pz) \\ Z &= z \end{aligned}\right\} \quad . \quad . \quad . \quad (25)$$

These co-ordinates are suitable to represent a guide of rectangular cross-section uniformly twisted about the Z axis. For constant Z the XY axes are obtained from the xy axes simply by a rotation through an angle pZ. Similarly the unit vectors i_X and i_Y are related to i_x and i_y by a rotation. The unit vector i_Z, however, differs from i_z owing to the fact that a Z translation, for fixed values of X and Y, rotates the radius vector to the point under consideration. It is readily shown that

$$i_Z = (i_z - pyi_x + pxi_y)M^{-1} \quad . \quad . \quad . \quad . \quad (26)$$

where $M^2 = 1 + p^2(x^2 + y^2)$.

Thus i_Z is not parallel to Z, but lies along a spiral path. It is not orthogonal to i_X and i_Y. However, by resolution along the z axis it is readily shown that the z and Z components of any vector A are related by

$$A_Z = MA_z \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (27)$$

where M is given by eqn. (26).

The requirements for an electric field E in the twisted guide are that $E_X = 0$ at $Y = \pm\frac{1}{2}a$, $E_Y = 0$ at $X = \pm\frac{1}{2}b$ and $E_Z = 0$ at the surface of the guide. When these conditions are satisfied it is readily seen that the normal component of H is also zero at the guide surface. Let ψ be a solution of

$$\operatorname{div}\operatorname{grad}\psi + k^2\psi = 0 \quad . \quad . \quad . \quad . \quad (28)$$

Then a solution of the wave equation[5]

$$E = \operatorname{curl}(i_z\psi),\ H = \left(\frac{j}{\eta k}\right)[\operatorname{grad}(\partial\psi/\partial z) + k^2 i_z\psi]$$

gives $E_x = \partial\psi/\partial y$, $E_y = -\partial\psi/\partial x$, $E_z = 0$ or, changing to XYZ co-ordinates,

$$E_X = \partial\psi/\partial Y,\ E_Y = -\partial\psi/\partial X,\ E_Z = 0 \quad . \quad . \quad (29)$$

Hence we need a solution of eqn. (28) subject to the conditions $\partial\psi/\partial Y = 0$ at $Y = \pm\frac{1}{2}a$, and $\partial\psi/\partial X = 0$ at $X = \pm\frac{1}{2}b$. The vanishing of E_Z at the guide surface is satisfied, since eqn. (29) has a zero Z component.

The form of eqn. (28) in helical co-ordinates may be found either from general considerations, or directly from the co-ordinate relations of eqn. (25). The details will not be given here. The equation for ψ is found to be

$$\begin{aligned} &(1 + p^2Y^2)\frac{\partial^2\psi}{\partial X^2} + (1 + p^2X^2)\frac{\partial^2\psi}{\partial Y^2} + \frac{\partial^2\psi}{\partial Z^2} + k^2\psi \\ &\quad + 2p\left(Y\frac{\partial^2\psi}{\partial X\partial Z} - X\frac{\partial^2\psi}{\partial Y\partial Z}\right) \\ &\quad\quad - p^2\left(X\frac{\partial\psi}{\partial X} + Y\frac{\partial\psi}{\partial Y} + 2XY\frac{\partial^2\psi}{\partial X\partial Y}\right) = 0 \quad . \quad (30) \end{aligned}$$

We seek a solution which varies as $\varepsilon^{-j\gamma Z}$, where γ is to be determined. An expansion in powers of $p = 2\pi/L$ is suggested, and we insert in eqn. (30)

$$\left.\begin{aligned} \psi &= (\psi_0 + p\psi_1 + p^2\psi_2 + \ldots)\varepsilon^{-j\gamma Z} \\ k^2 - \gamma^2 &= (\pi^2/a^2)(A_0 + A_1 p + A_2 p^2 + \ldots) \end{aligned}\right\} \quad . \quad (31)$$

Inserting eqn. (31) into eqn. (30) and equating powers of p leads to a set of equations for the ψ's. The zero-order equation is

$$\frac{\partial^2\psi_0}{\partial X^2} + \frac{\partial^2\psi_0}{\partial Y^2} + \frac{\pi^2}{a^2}A_0\psi_0 = 0 \quad . \quad . \quad . \quad (32)$$

Subject to the boundary conditions, and if only the dominant mode can propagate as will be assumed, the solution of eqn. (32) is

$$\psi_0 = \sin(\pi Y/a) \quad . \quad . \quad . \quad . \quad . \quad (33)$$

Hence from eqn. (32) $A_0 = 1$, as required to give $\gamma = k'$ for an untwisted guide ($p = 0$). The arbitrary multiplying constant which could appear in eqn. (33) has been taken as unity. As discussed in the previous Sections, in order to treat a unique solution, the form for the ψ's will not involve ψ_0 explicitly. The equation for ψ_1 is

$$\begin{aligned} &\frac{\partial^2\psi_1}{\partial X^2} + \frac{\partial^2\psi_1}{\partial Y^2} + \frac{\pi^2}{a^2}\psi_1 + \frac{\pi^2}{a^2}A_1\sin(\pi Y/a) \\ &\qquad + 2jk'(\pi X/a)\cos(\pi Y/a) = 0 \quad . \quad (34) \end{aligned}$$

The most general form fitting the boundary conditions is a double Fourier series

$$\psi_1 = \sum_0^\infty\sum_0^\infty A_{mn}\cos\left[\frac{n\pi}{a}(Y - \tfrac{1}{2}a)\right]\cos\left[\frac{m\pi}{b}(X - \tfrac{1}{2}b)\right] \quad . \quad (35)$$

Insertion into eqn. (34) gives

$$\begin{aligned} &\sum_0^\infty\sum_0^\infty A_{mn} \\ &\quad \left(\frac{m^2\pi^2}{b^2} + \frac{n^2\pi^2}{a^2} - \frac{\pi^2}{a^2}\right)\cos\left[\frac{n\pi}{a}(Y - \tfrac{1}{2}a)\right]\cos\left[\frac{m\pi}{b}(X - \tfrac{1}{2}b)\right] \\ &\qquad = \frac{\pi^2}{a^2}A_1\sin(\pi Y/a) + 2jk'(\pi X/a)\cos(\pi Y/a) \quad . \quad (36) \end{aligned}$$

On the left the coefficient of $\psi_0 = \sin(\pi Y/a)$ vanishes ($m = 0$, $n = 1$). Hence, by the usual process for determining the coefficients of a Fourier series, we find, for this term

$$0 = \int_{-\frac{1}{2}a}^{\frac{1}{2}a}\int_{-\frac{1}{2}b}^{\frac{1}{2}b}\sin\left(\frac{\pi Y}{a}\right)\left[\frac{\pi^2}{a^2}A_1\sin\left(\frac{\pi Y}{a}\right) + 2jk'(\pi X/a)\cos\left(\frac{Y\pi}{a}\right)\right]dXdY$$

This gives $A_1 = 0$, showing that, as anticipated, there is no first-order correction term in eqn. (31) for γ.

The remaining coefficients, A_{mn}, may be found from eqn. (36) in the usual way. On carrying out the necessary integrations and substituting back in eqn. (35), the following form for ψ_1 is found:

$$\psi_1 = \sum_{n=0}^{\infty}\sum_{m=1}^{\infty} \frac{4jk'b}{\pi^2 a}\,\frac{2-\varepsilon_n}{\dfrac{m^2\pi^2}{b^2}+(n^2-1)\dfrac{\pi^2}{a^2}}\,\frac{1-\cos m\pi}{m^2}$$

$$\frac{1+\cos n\pi}{1-n^2}\cos\left[\frac{m\pi}{b}(X-\tfrac{1}{2}b)\right]\cos\left[\frac{n\pi}{a}(Y-\tfrac{1}{2}a)\right] \quad . \quad (37)$$

where $\epsilon_0 = 1,\ \epsilon_n = 0\ (n > 0)$

The equation for ψ_2 may be shown to be

$$\frac{\partial^2\psi_2}{\partial X^2}+\frac{\partial^2\psi_2}{\partial Y^2}+\frac{\pi^2}{a^2}\psi_2-(\pi^2X^2/a^2)\sin(\pi Y/a)$$
$$+A_2(\pi^2/a^2)\sin(\pi Y/a)-(\pi Y/a)\cos(\pi Y/a)$$
$$+2jk'(X\partial\psi_1/\partial Y - Y\partial\psi_1/\partial X) = 0 \quad . \quad (38)$$

As with the determination of ψ_1, ψ_2 is assumed to be a double Fourier series, and after substitution into eqn. (38) the coefficient of $\sin(\pi Y/a)$ on the left-hand side is zero. Hence we have the equation from which A_2 may be found:

$$0 = \int_{-\frac{1}{2}a}^{\frac{1}{2}a}\int_{-\frac{1}{2}b}^{\frac{1}{2}b}\sin(\pi Y/a)[(-\pi^2X^2/a^2)\sin(\pi Y/a)$$
$$+A_2(\pi^2/a^2)\sin(\pi Y/a)-(\pi Y/a)\cos(\pi Y/a)$$
$$+2jk'(X\partial\psi_1/\partial Y - Y\partial\psi_1/\partial X)]dXdY \quad . \quad (39)$$

It is necessary to insert eqn. (37) into eqn. (39) and carry out the indicated integrations. The process is somewhat tedious but quite straightforward; the details will be omitted here. A_2 is found to be given by

$$0 = \frac{\pi^2 b}{2a}(A_2 - b^2/12) - \frac{ab}{4} - \frac{32k'^2b}{\pi^4 a}S \quad . \quad . \quad (40)$$

where

$$S = \sum_{n=0}^{\infty}{}' \sum_{m=1}^{\infty} \frac{2-\varepsilon_n}{m^4(n^2-1)^3}(1-\cos m\pi)(1+\cos n\pi)$$
$$\frac{b^2n^2(1-n^2)+a^2m^2(1+n^2)}{m^2\pi^2/b^2+(n^2-1)\pi^2/a^2} \quad . \quad (41)$$

The Σ' indicates that the $n = 1$ term is to be omitted.

The $n = 0$ term is

$$S_0 = \sum_{1,3\ldots}^{\infty}\left(-\frac{4a^2}{m^2}\,\frac{1}{m^2\pi^2/b^2-\pi^2/a^2}\right)$$
$$= -(4a^4/\pi^2)\sum_{1,3\ldots}^{\infty}\left(\frac{1}{m^2-b^2/a^2}-\frac{1}{m^2}\right)$$
$$= \tfrac{1}{2}a^4[1-(2a/\pi b)\tan(\pi b/2a)] \quad . \quad . \quad . \quad . \quad (42)$$

Similarly, by summing with respect to m, it may be shown that the rest of the series of eqn. (41) is

$$\sum_{2,4\ldots}^{\infty}\frac{a^2}{(n^2-1)^3}\left\{-\frac{n^2\pi^2b^2}{12}+a^2\frac{2n^2+1}{n^2-1}\right.$$
$$\left.\left[1-\frac{2a}{\pi b\sqrt{(n^2-1)}}\tanh\frac{\pi}{2}\sqrt{(n^2-1)}\frac{b}{a}\right]\right\} \quad . \quad (43)$$

Substituting into eqn. (40) gives A_2. From eqn. (31) γ may be found. Denoting the wavelength in the twisted guide by λ_T, so that $\gamma = 2\pi/\lambda_T$, it is found that

$$\frac{1}{\lambda_T^2} = \frac{1}{\lambda_g^2}\left\{1-\frac{1}{L^2}\left[\lambda_g^2\left(\frac{1}{2}+\frac{\pi^2b^2}{12a^2}\right)-\frac{\pi^2b^2}{3}+\frac{256a^2}{\pi^2}F(\phi)\right]\right\} \quad . \quad (44)$$

where $\phi = \frac{1}{2}\pi b/a$, and

$$F(\phi) = \tfrac{1}{2}(1-\tan\phi/\phi)+\sum_{2,4\ldots}^{\infty}\frac{2n^2+1}{(n^2-1)^4}\left\{1-\frac{\tanh[\sqrt{(n^2-1)}\phi]}{\sqrt{(n^2-1)}\phi}\right\}$$

The series $\sum_{2,4\ldots}^{\infty}\frac{2n^2+1}{(n^2-1)^4}$ is summable to

$$(\pi^4/16+3\pi^2/8-8)/16$$

The term $\frac{2n^2+1}{(n^2-1)^4}\,\frac{\tanh[\sqrt{(n^2-1)}\phi]}{\sqrt{(n^2-1)}\phi}$,

for values of n greater than about 2 approaches the form

$$\frac{2n^2(1+1/2n^2)}{(n^2-1)^4 n(1-1/2n^2)\phi} \simeq \frac{2n(1+1/n^2)}{(n^2-1)^4\phi} \simeq \frac{2n^3}{(n^2-1)^5\phi}$$

in which form it is summable in terms of the ζ function to

$$\frac{1}{256\phi}[10-21\zeta(3)+31\zeta(5)] = 16{\cdot}902/256\phi$$

Using these values an alternative form of eqn. (44) is found:

$$\frac{1}{\lambda_T^2} = \frac{1}{\lambda_g^2} - \frac{1}{L^2}\left\{(\tfrac{1}{2}+\phi^2/3)+\left(\frac{a^2}{\lambda^2}-\frac{1}{4}\right)\left[6+\pi^2-4\phi^2/3+\frac{G(\phi)}{\pi^2\phi}\right]\right\} \quad (45)$$

$$G(\phi) = -16{\cdot}902 - 128\tan\phi +$$
$$4\sum_{1}^{\infty}\frac{1}{(n^2-\frac{1}{4})^5}\left\{n^3-(n^2+\tfrac{1}{8})\sqrt{(n^2-\tfrac{1}{4})}\tanh[2\phi\sqrt{(n^2-\tfrac{1}{4})}]\right\}$$

The series is extremely rapidly convergent; for all practical purposes the first term suffices.

The function $6+\pi^2-4\phi^2/3+G(\phi)/\pi^2\phi$ is plotted against $\phi = \frac{1}{2}\pi b/a$ in Fig. 6. Over the usual guide range of $0{\cdot}3 < b/a < 0{\cdot}6$ there is a substantial cancellation amongst the terms, and their resultant is small. However, as $b \rightarrow a$ the $\tan\phi$ term increases, becoming infinite for a square guide. For a narrow rectangular guide the polarization of the dominant mode is twisted as the guide twists. As the guide becomes more

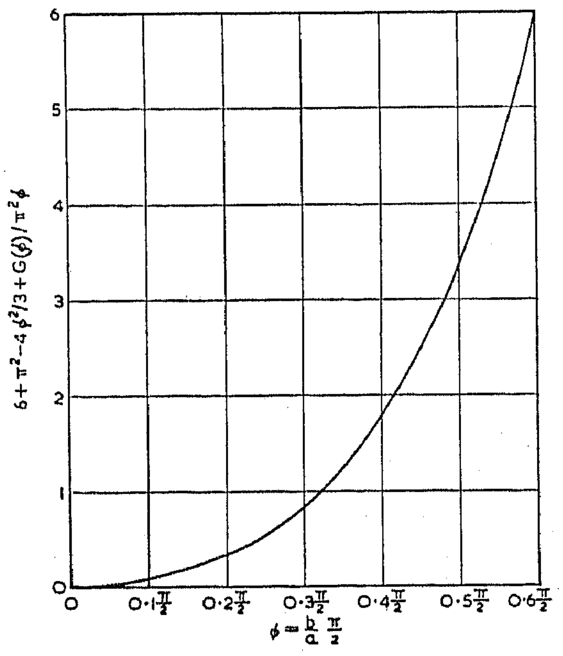


Fig. 6.—Plot of correction formula against ϕ for twisted guide.

and more square in shape, there is a tendency for the cross-polarized mode to be excited. This may be seen from the $n = 0$, $m = 1$ term of eqn. (37). The square guide is degenerate in this respect, and the solution cannot take the form of a single mode whose polarization twists with the guide. Since this assumption is really implicit in the choice $\psi_0 = \sin(\pi Y/a)$, this is the cause of the breakdown of eqn. (45) at $b = a$. Two coupled modes would apparently be necessary to solve the problem by the above process, but an attempt to do this has not so far met with success.

Fig. 7 shows the correction to the straight-guide wavelength for a series of values of λ/a and of $\phi = \frac{1}{2}\pi b/a$. The quantity e is defined by

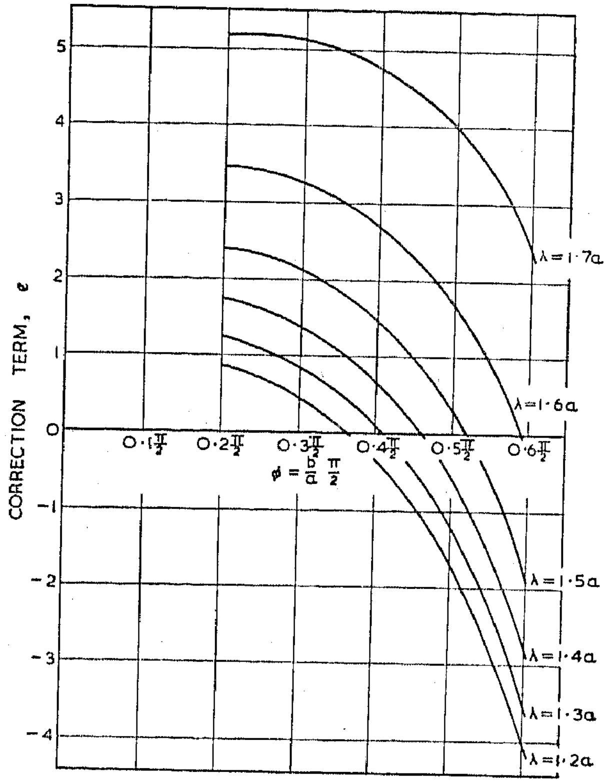


Fig. 7.—Plot of correction coefficient against ϕ for twisted guide.

$$\frac{1}{\lambda_T^2} = \frac{1}{\lambda_g^2}(1 - ea^2/L^2) \quad . \quad . \quad . \quad . \quad . \quad (46)$$

and is plotted against ϕ in the diagram. The correction term is seen to vanish at some wavelength in the usual band, but the correction is far from negligible except for twists many times longer than the guide width.

(6) CONCLUSIONS

The wave equation has been solved for three configurations of rectangular-cross-section guide by the utilization, as independent variables, of an axial co-ordinate and two cross-section co-ordinates. The form of solution has been an expansion in terms of a parameter which represents the departure of the system from the straight rectangular guide. In all cases the propagation coefficient has involved the square of the expansion parameter, so that two steps in the expansion of the field give one term in the expansion of the propagation coefficient. The expansions have been completed for the first two perturbation terms in the field and the first term in the propagation coefficient.

For the circular bends the results are presented in eqns. (12) and (24) for the guide wavelength, the correction terms being plotted in Figs. 2 and 4. Within the usual operating band both correction terms go through zero, and are rather small, numerically, within the range. The expansion process is straightforward, and can be readily continued, if required, to further terms.

For the twisted guide, the results are presented in eqns. (45) and (46), with numerical results in Figs. 6 and 7. Although the correction terms go through zero within the usual operating band, the terms are far from being numerically small; in fact, as the guide cross-section approaches a square shape the correction becomes very large, showing the inadequacy of the expansion in this region. Since the expansion process involves a double summation per step, the next term in the propagation coefficient would involve a six-fold summation, from which it is apparent that the method of solution is unsatisfactory when further terms are required. The problem of the twisted square guide remains unsolved.

(7) ACKNOWLEDGMENTS

The author is obliged to Standard Telecommunication Laboratories, Ltd., for facilities provided for the preparation of the paper, and for permission to publish it.

(8) REFERENCES

(1) Jouguet, M.: "Les effets de la courbure sur la propagation des ondes électromagnétiques dans les guides à section circulaire," *Câbles et Transmission*, July, 1947, p. 133.
(2) Lewin, L.: "Advanced Theory of Waveguides" (Iliffe and Sons), p. 37.
(3) Marcuvitz, N.: "Waveguide Handbook" (McGraw-Hill), p. 334.
(4) Marcuvitz, N.: *loc. cit.*, p. 335.
(5) Lewin, L.: *loc. cit.*, p. 12.

621.385.16.029.6

Paper No. 1760 R
Jan. 1955

NOISE IN CUT-OFF MAGNETRONS

By R. C. GLASS, B.A., B.Sc., G. D. SIMS, M.Sc., B.Sc., and A. G. STAINSBY, B.Sc.

(*The paper was first received* 31*st May, and in revised form* 30*th August*, 1954.)

SUMMARY

Experiments have been made on the characteristics of magnetron oscillators at anode voltages from zero to the value at which the oscillations start. In particular, the noise in the anode current was investigated. Discrete peaks of noise were found in the spectrum, and the frequency at which they occurred varied in a regular manner with anode voltage and magnetic field strength. A qualitative explanation of these peaks is given, and the implication of these results on the mechanism of build-up of oscillations and the selection of modes of oscillation is discussed.

(1) INTRODUCTION

The theory of the magnetron oscillator so far published has in the main dealt with the conditions from which oscillations can build up, on the assumption that a high level of noise or a low-level injected signal is already present. Hartree and Bunemann,[1] in their analysis of the interaction between electrons in the anode-cathode space and a rotating r.f. wave on the anode, obtain an expression for the minimum potential required to enable electrons to reach the anode when the wave has a vanishingly small amplitude. This minimum voltage, or threshold voltage as it is usually called, depends on the number of repeats of the r.f. wave around the anode, the magnetic field and the geometry of the interaction space. It is found in practice that the threshold voltage agrees closely with the voltage at which the magnetron starts to oscillate. However, to account for the high rate at which oscillations build up, it is also necessary to assume a high level of noise.

Several papers have appeared which analyse the space charge based on a particular model. Of these, Gabor[2] considers the electron cloud as completely disordered and investigates the effect of initial velocities on the cloud characteristics. Hok[3] follows a similar analysis, but neither in this case, nor that above, is the fundamental disordering process fully discussed. Harris[4] investigates the stability of the single-stream steady state and finds that the electron cloud contains frequency components other than random ones, but no attempt is made to correlate the theoretical results with experiments. Harris initially interpreted these frequencies as instabilities, but later reversed this opinion.

It is well known, to the regret of many students who have used this method in an attempt to measure the ratio of charge to mass of an electron, that as the anode voltage of a cylindrical magnetron is increased, the valve starts to take current at a much lower voltage than the critical value at which the electron paths should just touch the anode.

This has been variously attributed to mechanical or electrical irregularities of the system, to electronic oscillations, space-charge amplification or electron collisions. Other phenomena, such as cathode back-bombardment and enhanced noise, also await satisfactory explanation.

On the theoretical side the equations governing the steady-state motions of the electrons yield two possible types of solution. In one case the electron motion has a double-stream nature, in which all electrons leaving the cathode ultimately return to it, and thus at each point in space there will be electrons with components of radial velocity directed both towards and away from the cathode. The other solution is the single-stream steady state in which the electrons move in concentric circles about the cathode with zero radial velocity. Although there are objections to both these solutions, there has as yet been no experimental evidence which would enable a decision to be made between them.

It is evident, therefore, that a determined experimental study of the behaviour of the magnetron space-charge cloud is long overdue, both to investigate the anomalies mentioned above and to decide between the various theories relating to the pre-oscillating magnetron which have been put forward.

It seemed that two items particularly merited attention:

(*a*) An experimental study of the magnetron space-charge cloud in the pre-oscillating state, in order to determine its exact role in the build up of oscillations, and from these measurements to decide, if possible, whether a single- or double-stream state is present.

(*b*) An attempt to find the reason for the high electron temperature and to see whether this was a suitable term to use in describing the state of the electrons, since from theoretical considerations it was thought that the distribution of energy with frequency should not be continuous but should contain one or more peaks.

For these reasons an investigation was first made of the noise output as measured by fluctuations in the anode current and as picked up either by a probe or by the cathode supports of the magnetron itself. Preliminary measurements soon showed that the noise did not have a uniform distribution, but contained some well-defined maxima.

Measurements of the noise at a given frequency showed one or more peaks as the anode voltage was increased from zero to the threshold of oscillation. Further investigation showed that the voltages at which the noise peaks occurred varied with frequency. It became evident that at any voltage below the threshold of oscillation there was present a general background of noise which increased in amplitude as the anode voltage was increased, and superimposed on this background were more or less clearly defined peaks. These have been found at frequencies from 0·5 Mc/s up to, in some cases, 4 000 Mc/s.

The paper is concerned with these peaks and their origin.

(2) MEASUREMENT OF NOISE

When the experiments were started it was thought that it would be necessary to make a magnetron having a probe in the interaction space in order to measure the noise. A valve of this type was, in fact, constructed. It soon became clear, however, that many useful results could be obtained by measuring the noise generated in an ordinary c.w. magnetron, by measuring either the fluctuations of anode current or the noise power radiated by the heater leads. This technique was useful because well-defined peaks of noise are present and a relation could be found between the voltage and frequency at which they occur. It is not possible by such a simple method to obtain a relation between the absolute noise output and frequency for any given anode voltage. This could only be done

---

Written contributions on papers published without being read at meetings are invited for consideration with a view to publication.

The paper is a communication from the Staff of the Research Laboratories of The General Electric Company, Limited, Wembley, England.

Mr. Sims is at the Atomic Energy Research Establishment.

by using a non-frequency-selective probe, and such measurements present great technical difficulties.

Measurements have been made on several multi-segment magnetrons having 8, 10 and 18 segments. All these valves were designed to oscillate at a frequency of about 3 000 Mc/s in the π-mode. However, as noise peaks were measured from 0·5 Mc/s up to and beyond the π-mode frequency, it is evident that for most frequencies up to, say, 0·8–0·9 of the π-mode frequency the presence of the segments had no effect and the valve could be treated as a cylindrical diode.

For frequencies up to 100 Mc/s, measurements were made of the anode-current fluctuations in a 50-ohm resistance in series with the magnetron. The relative noise power was measured in the conventional way using an oscillograph and a square-wave signal generator as a source of reference voltage.[5] This method is preferable to one using a square-law detector and meter since it enables one to distinguish between fluctuation noise and hum or coherent oscillations. At this point it should be stated that what is described as "noise" is only so called because of its appearance on the oscillograph screen. No attempt has been made to check the randomness. All that can be said is that as displayed on an oscillograph and with the bandwidths used (about 1 Mc/s maximum) it is indistinguishable from random noise.

For higher-frequency measurement a capacitive connection was made to the cathode lead of the magnetron and fed into a crystal mixer excited by a variable-frequency oscillator. A triode was used for frequencies up to 3 000 Mc/s, and klystrons were used for frequencies above this. The output from the mixer was fed into a 13-Mc/s i.f. amplifier of 4 Mc/s bandwidth, which in turn fed a video amplifier and oscillograph. A signal generator at 13 Mc/s which could be modulated with square waves was used as a reference voltage (see Fig. 1).

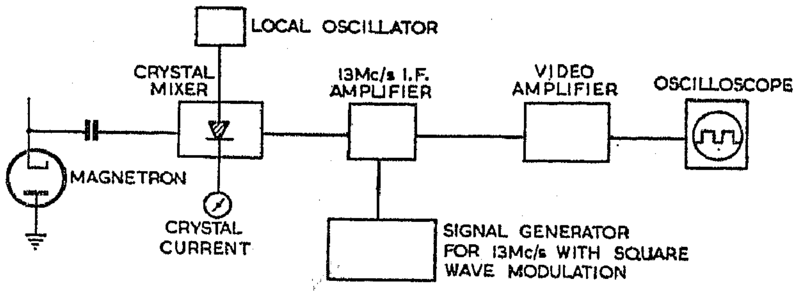


Fig. 1.—Experimental arrangement used to measure magnetron noise.

In making these measurements the current in the first detector was fixed and the height of the oscillograph picture was kept constant by varying the gain of the i.f. amplifier. This ensured that the first and second detectors were always working at the same level. A check was made of the variation in bandwidth of the receiver with gain, and it was found to be negligible. Also, since the noise level of the receiver was usually small compared with that of the magnetron, it was found that a change in noise of 1 dB could be measured.

In this receiver system, since there was no method for image-frequency rejection over so wide a frequency range, the output represents the sum of the noise from two bands 26 Mc/s apart in frequency. Thus if a sharp peak of noise were received, it would appear as a double signal. By the use of a transmission cavity wavemeter in cascade with the lead from the magnetron the correct frequency at which the peak occurred could be distinguished.

(3) RESULTS

(3.1) Effects of Anode Potential and Magnetic Field

Measurements were made of the noise output with variation of voltage for constant magnetic field and with the receiver at one frequency. The voltage was increased slowly from zero up to the value at which the valve oscillated. Typical curves for two frequencies are shown in Fig. 2. The noise is plotted relative to an arbitrary level which is not the same for the two curves. This must necessarily be the case since no attempt was made to obtain a non-frequency-selective coupling to the space charge. It is seen that, in general, the noise increased as the anode voltage was increased up to the value at which the valve started to oscillate. At intervals there were also peaks, which increased in amplitude with higher voltages, and there was always a maximum near the oscillation voltage. If curves such as these were plotted for several frequencies it was found that a given peak moved to higher frequencies as the anode voltage was increased. The relative heights of the peaks also varied, and at some frequencies no peaks could be found at all. It is not known whether this is a property of the magnetron or whether it is a function of the method of measurement, due, for example, to some frequency selectivity. The latter is not thought probable. As measurements were taken to higher frequencies the peaks became less pronounced. With the 18-segment valve (a valve with a large anode and cathode diameter) no peaks were found for frequencies above 2 500 Mc/s. The curve of noise against voltage then became smooth, rising to a maximum at the oscillation voltage. However, measurements on the 10-segment valve showed the presence of peaks up to and beyond 3 500 Mc/s, while with the 8-segment valve peaks could be found up to 4 700 Mc/s.

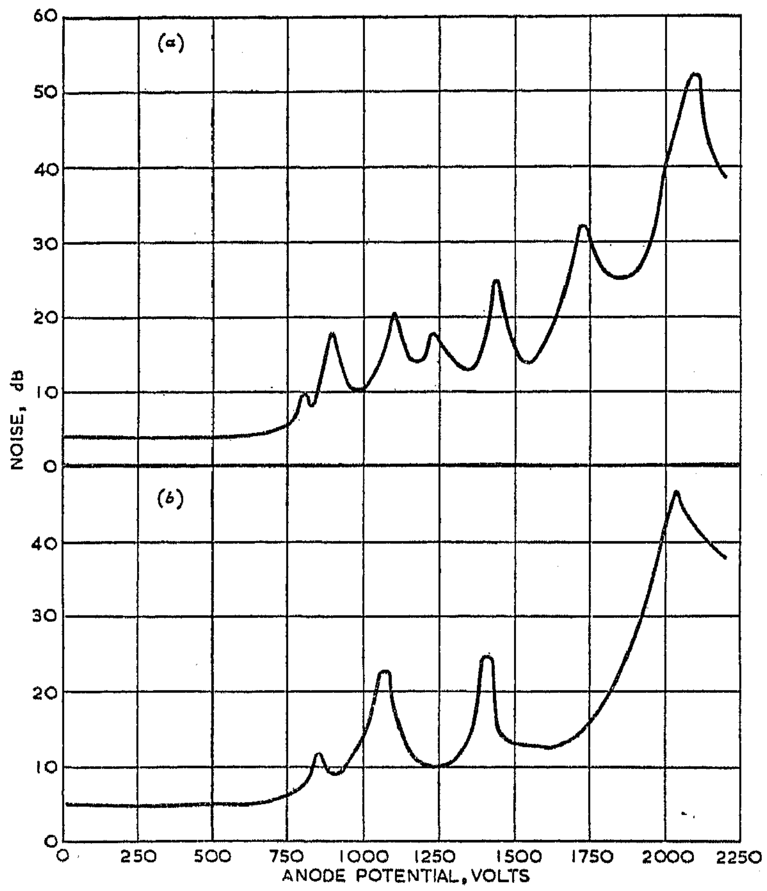


Fig. 2.—Typical noise/voltage curves for 10-segment valve ($H = 2\ 110$ oersteds, anode diameter = 0·635 cm, cathode diameter = 0·325 cm).

(a) 2 200 Mc/s.
(b) 1 000 Mc/s.

The voltages at which the noise peaks occurred were plotted against frequency, and it was found that the peaks lay on a series of curves diverging apparently from the origin (see Fig. 3). Since it was thought that the maxima occurred at voltages corresponding to an n-fold azimuthal variation in the space charge, it was natural to ascribe to these lines the numbers $n = 1$, 2, 3, 4, etc. On examining the curves to find such relations, it

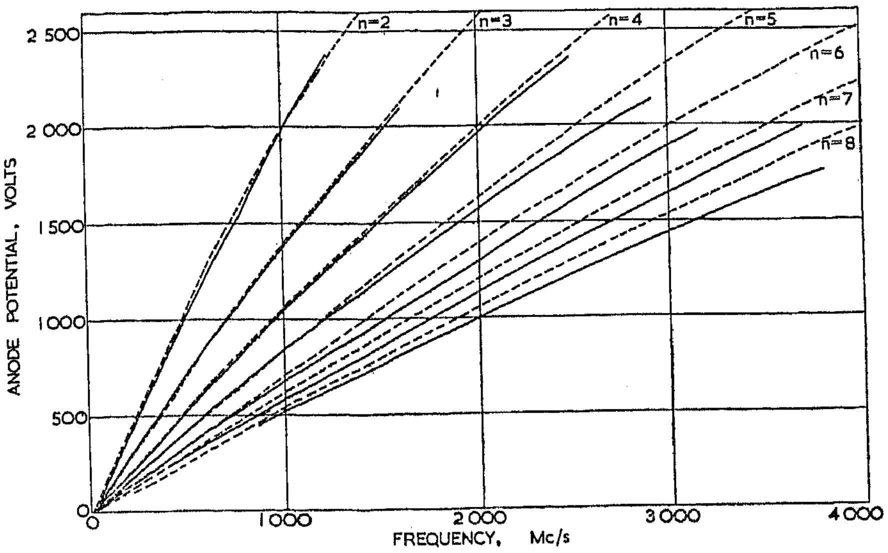


Fig. 3.—Peak-voltage/frequency curves for 10-segment valve ($H = 2\ 110$ oersteds, anode diameter = 0·635 cm, cathode diameter = 0·325 cm).

— — — Threshold lines.
——— Experimental lines.

was found that with the 18-segment valve the lines for $n > 2$ of the Hartree–Bunemann threshold relation fitted the curves very well. These equations take the form

$$V = AH\left(\frac{f}{n}\right) + B\left(\frac{f}{n}\right)^2 \quad . \quad . \quad . \quad . \quad (1)$$

where A and B are geometric constants for the magnetron, H is the magnetic field strength, f is the frequency and n is the order of the mode under consideration.

There was, of course, no external means by which the order to which each curve corresponded could be established. Further work on different sizes of valve led to the abandonment of the idea that the experimental lines were in fact the threshold lines, particularly measurements at low n numbers, where two lines have been found at voltages greater than the threshold voltage for $n = 2$ and neither of these lines fits the $n = 1$ threshold curve.

A series of measurements was also made for different magnetic field strengths, keeping the frequency constant. Again a series of lines was found—apparently straight within the limits of experimental error. Typical results of these measurements are plotted in Fig. 4. Again there is a similarity to the threshold-voltage lines but not complete coincidence.

In this case it was possible to vary the anode voltage and magnetic field simultaneously, keeping the noise peak in view on the oscillograph, and thus to plot the variation of peak voltage with field directly for any particular peak.

(3.2) Cathode Temperature

Measurements were made of the height of a noise peak at a fixed frequency with variation in cathode temperature. The cathode temperature was measured through a window in the side of the valve by means of a pyrometer. Measurements were also made of the emission at the knee of the emission curve. In this condition the cathode (thorium-coated tantalum) was practically saturated. The noise and saturated emission are plotted in Fig. 5. It is seen that the noise increases by about 12 dB as the cathode temperature is increased from 1 200° C to 1 800° C. The lowest temperature was that at which there was just sufficient emission to sustain oscillation, and the maximum was close to the melting point of thorium.

(3.3) Back-Bombardment

The back-bombardment of the cathode is of interest, since any excess energy which is given to the electrons must be derived from some interaction in the anode-cathode space, and ulti-

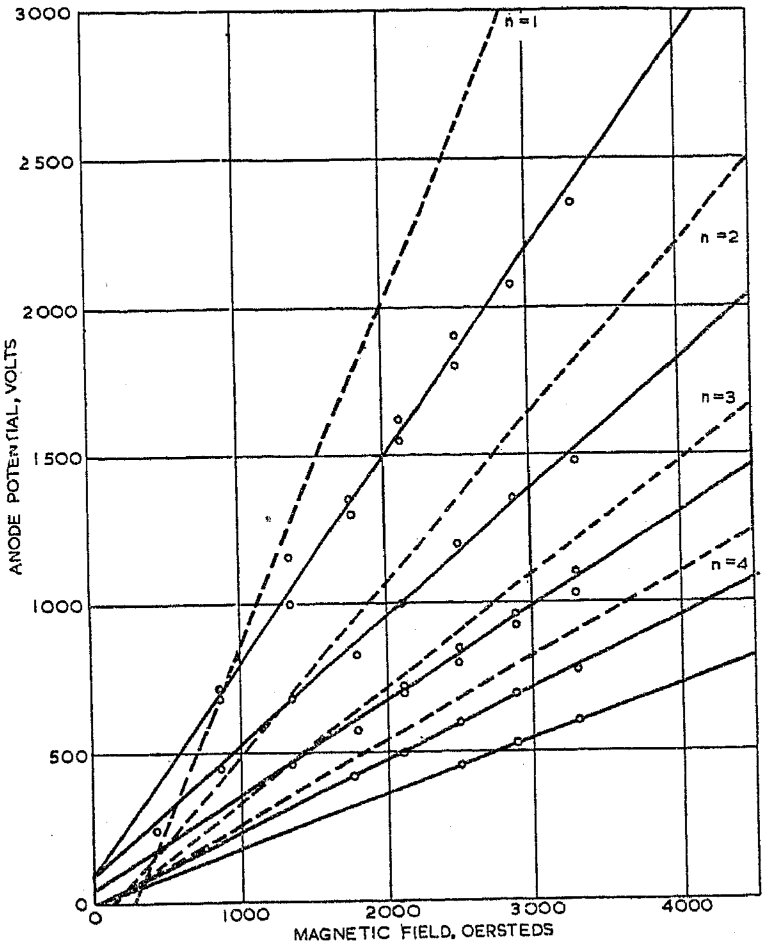


Fig. 4.—Peak-voltage/magnetic-field plot for 10-segment valve ($f = 486$ Mc/s, anode diameter = 0·635 cm, cathode diameter = 0·325 cm).

- - - - Threshold lines.
——— Experimental lines.

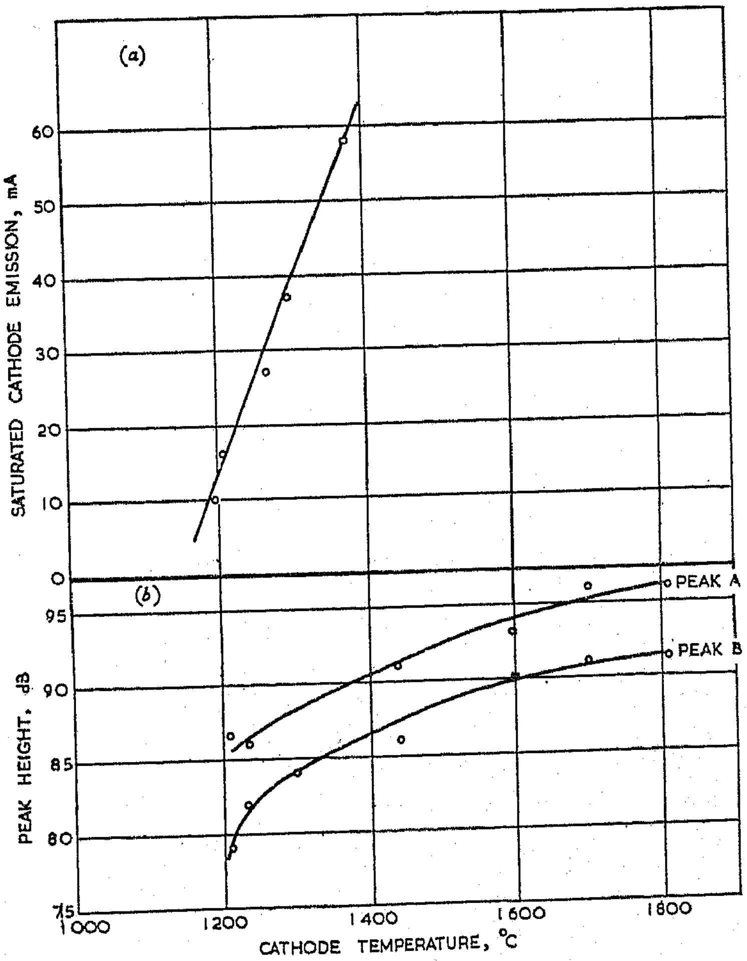


Fig. 5.—(a) Variation of saturated cathode emission with temperature.
(b) Variation of peak height with cathode temperature.
$H = 900$ oersteds.
$f = 80$ Mc/s.

mately from the d.c. power supply. Accordingly the back-bombardment power was measured in relation to the anode current and anode voltage. Measurements were made using a photocell and reducing the power in the heater until the indication was equal to that obtained before the h.t. supply was applied. This method assumes that the temperature distribution of the cathode due to the heater is the same as that due to back-bombardment. This is not strictly true, but the results are qualitatively valid. In Fig. 6 the back-bombardment, and back-bombardment as a percentage of the power into the valve, are plotted against anode voltage. It is seen that the back-bombardment increases steadily up to the voltage at which oscillations commence (1·85 kV), where it suddenly falls, indicating that more power is extracted from the electrons for the production of true oscillations. The curve of percentage back-bombardment falls steadily from a value of 60% and falls very rapidly from 30 to 20% as true oscillations build up.

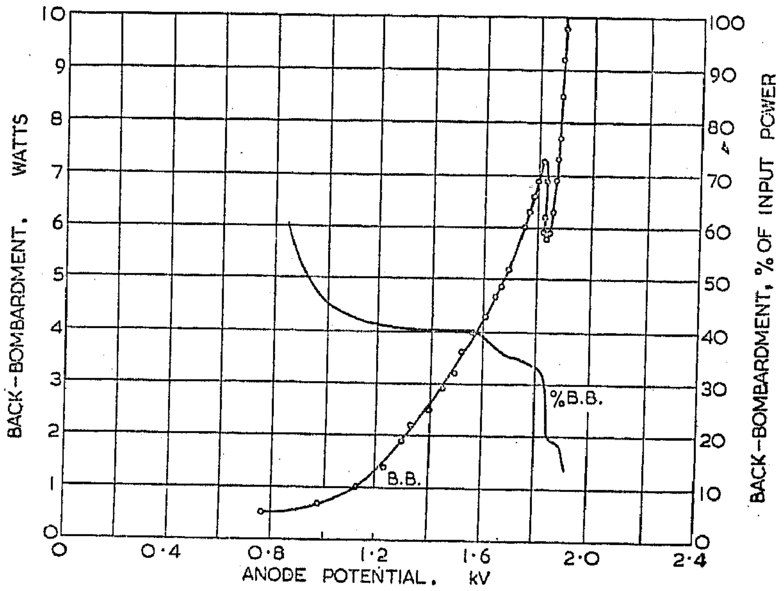


Fig 6.—Back-bombardment and percentage back-bombardment as a function of anode potential ($H = 900$ oersteds, 18-segment-valve, anode diameter = 1·25 cm, cathode diameter = 0·7 cm).

In general in an oscillating valve the percentage back-bombardment falls to 3–5% by the time the normal operating point is reached.

(4) THEORETICAL CONSIDERATIONS

A full discussion of the theory of the pre-oscillating state will not be attempted here. However, a brief summary of some theoretical considerations which may be of importance will be presented, together with a discussion of the qualitative picture to which they lead.

The observed peaks in the noise output may be due either to a selective amplification of noise within the space-charge cloud or to independent oscillation thereof. The meaning of these two terms requires some clarification in order to see how they may be applied to the magnetron. In a linear device with a clearly defined input and output, amplification may be said to occur if a signal applied at the input increases in strength along the length of the tube. If, on the other hand, in a similar tube the signal strength at any point increases with time to a first-order approximation, the system is unstable and is inherently oscillatory. Some feedback of energy will be required for this to occur.

In the magnetron there are two possible feedback mechanisms. If a double-stream state is admitted, electrons from the outer edge of the space-charge cloud can feed back energy towards the cathode, which is the origin of the noise. Alternatively, if pure amplification is postulated, the feedback can occur around the interaction space, and assuming some phase coherence azimuthally, oscillation will take place. Guénard and Huber[6] have assumed pure amplification to explain the mechanism of a linear crossed-field device, and they have extended this to explain some oscillations in cylindrical magnetrons with filamentary cathodes. It seems likely, in fact, that both mechanisms of feedback are present in the magnetron.

Owing to the phase coherence around the anode which is required to make use of the amplification mechanism, we might suppose that the oscillations would be relatively sharp. However, since there are no resonant circuits to control them, it seems unlikely that they would appear as anything more than peaks in the noise spectrum, as is observed. At higher frequencies it is doubtful whether all the peaks present would be observed with the method of coupling used. In this region the anode block presents a low impedance to all frequencies displaced from those of its own spectrum.

In the foregoing, some of the characteristics which must apply to the interaction of electrons and waves in a magnetron-type structure are discussed. These will be equally valid for either a pure amplifying or oscillating process, or for a combination such as has been envisaged.

For energy interchange to occur between a wave and the rotating electron cloud, it is necessary that their angular velocities should be nearly equal, i.e. some relation of the form

$$\left|\frac{\omega}{n}\right| \simeq \left|\frac{\bar{v}_t}{r}\right| \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

should be obeyed. Here $\bar{v}_t$ is the mean tangential electron velocity, r some radius as yet unspecified, and n is the number of repeats of the wave pattern round the anode structure. This relation is the basis of operation of most high-frequency valves and accelerators, and in the case of the magnetron, represents the first approximation to the threshold relation,[7] particularly for low values of f/n. It seems reasonable, therefore, to try to correlate the points obtained with the threshold lines. As has already been shown, these lines show much similarity to those followed by the peaks, and the general trend is identical. However, the threshold lines are not by any means followed strictly, and reasons for this deviation must be found.

It was thought that some condition controlling the radial motion of the electrons might be of importance, and an investigation of this is to be published elsewhere. Consideration of radial motion implies that a double-stream model be adopted, and this is considered to be legitimate, since whatever are the current predictions of theory in this branch of the subject, it is clear, from consideration of the back-bombardment phenomenon alone, that a powerful returning stream does impinge on the cathode. It is, in fact, thought that the observed peaks in the noise spectrum arise from the same pre-oscillation phenomenon which is responsible for both anode current and back-bombardment.

The theoretical investigation referred to showed that there were certain frequencies for which no back-bombardment of the cathode occurred. That these frequencies could be present as oscillations, could be further inferred from the theory. It was clear from these experiments that the threshold relation is obeyed fairly closely, and it was thought that a close coincidence between these and the zero back-bombardment curves might exist. This may indeed be the case at higher frequencies, but there is no correspondence at the lower frequencies. The zero back-bombardment curves are compared with the relevant threshold curves for the 18-segment valve referred to in the experiments in Fig. 7, and they take the general form

$$\frac{\omega}{n} - \frac{n\bar{v}_t}{r} = \mathrm{f}(\tau) \quad . \quad . \quad . \quad . \quad . \quad (3)$$

where τ is the transit time of an electron across the electron cloud.

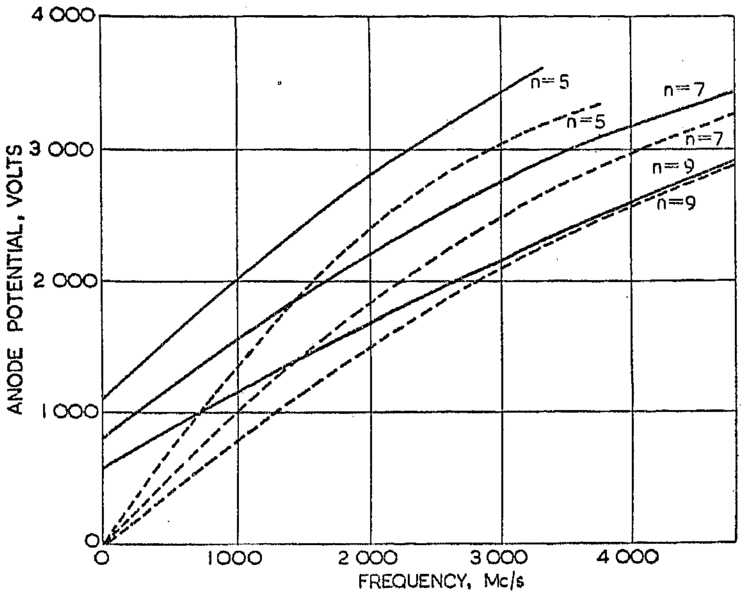


Fig. 7.—Comparison of threshold lines and zero back-bombardment lines ($H = 900$ oersteds, anode diameter $= 1\cdot25$ cm, cathode diameter $= 0\cdot7$ cm).

- - - - Threshold lines.
——— Zero back-bombardment lines.

The importance of this relation from the point of view of initiation of oscillations and mode change might well be considerable, for it is not unreasonable to expect that, when the zero back-bombardment lines cross the threshold lines, the pre-oscillation peaks undergo an increase in amplitude. If such a coincidence occurred for any n number, the valve might well commence oscillation in that mode. Similarly, if the hypothesis that the main cloud is little disturbed when the oscillation sets in be adopted (and there is much evidence to support this), the same considerations could lead to mode-changing. It is now clear, however, from the experimental results, that in determining the frequencies of oscillation, the threshold criterion is by far the more important relation. The cathode-damping condition may be that responsible for deviation from the threshold lines, although if this is so it does not seem to be practicable to calculate how much the deviation would be.

The actual detection of peaks will occur, in general, when a receiver couples to the electromagnetic field associated with them. At the lower frequencies, the peaks possess very little electromagnetic field and arise mainly from an interchange of electronic kinetic energy with the electric field. We must consider the waves existing at these frequencies, then, as being predominantly space-charge waves. In a magnetron, there will always be a coupling to the cathode supports, in the first place from the axial thermal velocities of the electrons, and in the second place from any inhomogeneity or slight skewness of the magnetic field (something of which always exists even under the most carefully controlled conditions). As the voltage is raised, it is observed that the general noise level increases. This is due, at least in part, to the fact that increasing the voltage increases the frequency and therefore the electromagnetic field, thus giving stronger coupling to the outside. This increase in coupling with increase of voltage is of the utmost importance practically, since, not only does it imply greater coupling to the resonant-cavity system, but in addition the amplitude of the electromagnetic field associated with the space charge at any particular frequency will have a strong influence on the mode selection.

The measured dependence on cathode temperature suggests that at least part of the process may be of the nature of a noise amplification, although, as suggested earlier, if increase in thermal velocity gives rise to increased couplings this will account for some of the apparent rise.

The reason for the disappearance of peaks above 2 500 Mc/s in the case of the 18-segment valve is very probably due to the weaker coupling of the higher-order modes to the cathode supports. It is possible, however, that the effect may be real and that pre-oscillations cannot build up for an n number greater than a certain value. However, this runs counter to general magnetron experience, where π-mode operation has been obtained with very large numbers of segments.

In general the percentage back-bombardment curves obtained are similar in form to those of Jepsen and Muller,[8] their shape being not inconsistent with the ideas expressed above. The curve shown in Fig. 6 is not an entirely typical one in this sense, although its general trend is much the same.

Although the curves of anode potential against frequency (see Fig. 3) appear to emanate from the origin, it is clear that an energy in excess of the thermal energy must be available before any amplifying or oscillation process of the magnitude of that observed can be set up. This energy is derived from the anode current alone, and at low frequencies it is dissipated almost entirely as energy of back-bombardment. As the voltage and frequency increase, more energy is stored and radiated by the electromagnetic field, and hence less is available for back-bombardment. This, of course, involves the assumption that those electrons which reach the anode have lost energy to the field—a fact clearly borne out by the relatively small thermal dissipation.

(5) THE PROCESS BY WHICH A MAGNETRON STARTS TO OSCILLATE

Let us consider the implication of these results on the build-up of oscillations in a multi-cavity magnetron. It is seen that, as soon as the electron cloud is excited to an energy above the thermal value, it commences to oscillate at one or more initially low frequencies with azimuthal variations of low order. As the anode voltage is increased and more energy is fed into the space charge, the frequencies increase and higher-order azimuthal variations become evident, the actual frequencies at which these resonances occur being determined by a relation of the same general form as eqn. (1). This relation will involve, amongst other things, the geometrical parameters of the block, and will be satisfied by the measured frequencies at which noise peaks occur. The curves on which the peaks lie are represented by the full lines in Fig. 8, where they are compared with the threshold lines. One of these curves will correspond most closely over some part of its range to the threshold-voltage line of a mode in the anode block having the same azimuthal periodicity. Thus, at the threshold voltage corresponding to the block resonance

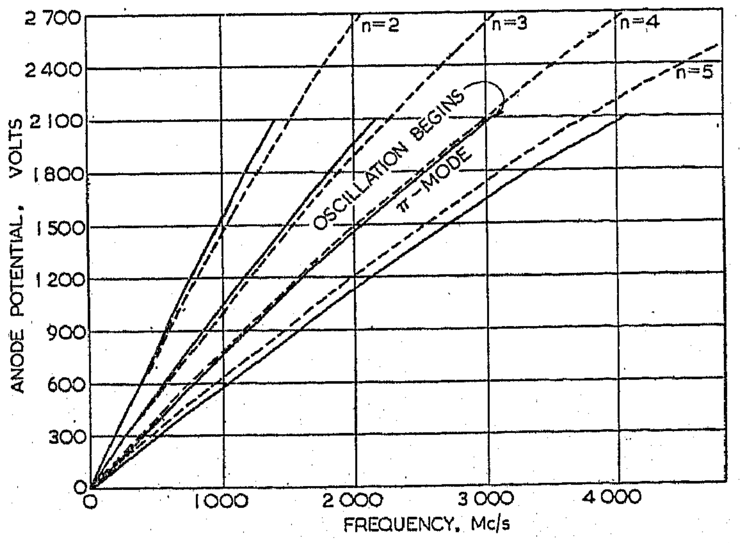


Fig. 8.—Comparison of experimental pre-oscillation lines with theoretical threshold lines for 8-segment valve ($H = 1\,944$ oersteds, anode diameter $= 0\cdot508$ cm, cathode diameter $= 0\cdot20$ cm).

- - - - Threshold lines.
——— Pre-oscillation lines.

this mode, there will already be present in the space charge a high-amplitude oscillation having the same periodicity as the block resonance. This mode, therefore, will receive a strong excitation and will tend to build up. It would be possible by choosing the appropriate mode separation to have another pre-oscillation coincident with another mode, in which case some other form of mode competition must decide which oscillation will build up. Once the build-up occurs the focusing forces due to the block fields will tend to take control. This simple excitation of the π-mode ($n = 4$) is shown in the curves obtained on an 8-segment strapped valve (see Fig. 8). This valve was a c.w. valve which would start to oscillate in the π-mode at 3 000 Mc/s. It is seen that the pre-oscillation curve coincides over a considerable length with the threshold curve, and that as the frequency is increased to 3 000 Mc/s, corresponding to the π-mode resonance, the valve bursts into oscillation at a voltage corresponding to the threshold voltage (within experimental error).

These results also indicate a second effect which may be important in the phenomenon of mode change. There is, at present, no satisfactory explanation of the factors which cause a magnetron to change its mode of operation from one to another. It is clear that the threshold voltage plays an important part, but it does not offer a complete explanation. It is essential for a mode change that the threshold voltage of the next mode shall be reached, otherwise the necessary energy conditions are not fulfilled. One might therefore expect that there will be a different effect if the pre-oscillation characteristic lies above or below the threshold voltage. If the pre-oscillation-voltage line is above the threshold-voltage line, the anode potential can be increased until the pre-oscillations excite sufficient voltage for the change to take place. However, if the pre-oscillation line is below the threshold line, the pre-oscillation will already have reached a higher frequency than that of the resonant mode by the time the voltage conditions are fulfilled. The amplitude in the new mode will thus be reduced and there should be less tendency for the change to take place. It seems quite probable that the pre-oscillations do, in fact, persist in the region of the cathode even when the valve is oscillating. In one experiment which tended to verify this, the noise output of a magnetron in resonant modes other than the π-mode was observed. The noise output was found to increase up to and beyond the voltage at which the π-mode was excited.

Consider now the build-up in the hypothetical case of an unstrapped 8-segment block in which the frequency of the $n = 3$ mode is more or less equal to that of the π-mode, ($n = 4$). As the voltage is increased the threshold for $n = 5$ (the reverse component of the $n = 3$ mode, having identical symmetry so far as the block is concerned) is reached first, then $n = 4$ and finally $n = 3$. On the knowledge of the threshold voltages alone, it would be expected that the $n = 5$ mode would build up first. In fact, it is found from experiment that this can be the case, but it is not necessarily so. By suitable adjustment of cathode diameter the π-mode may be the first to be excited. This has been explained in terms of the optimum coupling impedance being obtained for different ratios of cathode-to-anode diameter, but it may well be that by adjustment of cathode diameter the pre-oscillation line for the π-mode can be made to correspond to the threshold line, whilst that for the $n = 5$ and $n = 3$ modes does not.

The divergence of the lines is most marked for low n numbers, and thus for higher modes this method of mode selection becomes increasingly more critical. It is found, in fact, that in the design of magnetrons having large numbers of segments the cathode-to-anode diameter ratio has to be defined within very narrow limits.

(6) CONCLUSIONS

The experimental results described above have shown the presence of oscillations in the space-charge cloud down to extremely low voltages. Thus for all practical purposes the space charge can be considered as being always in a state of oscillation, and a true steady state therefore never really exists. The question of whether a single- or double-stream steady state is present is thus largely an academic one.

The oscillations in the space charge give an explanation of the high electron temperatures which have been observed. It would be possible to use the term "temperature" in describing the general background of noise between the peaks, but the peaks themselves are essentially noise-modulated oscillations.

The results obtained also throw some light on the behaviour of the magnetron space-charge cloud prior to the build-up of the main oscillation, and, in particular, give an explanation of the high rates of rise of r.f. amplitude which are achieved with pulsed magnetrons. It seems most probable that the pre-oscillation modes also play a large part in deciding in which mode a magnetron will start to oscillate. If the qualitative explanation given above of the role of these oscillations is correct, an investigation of the effect on them of various magnetron parameters (e.g. cathode diameter) would no doubt prove fruitful in throwing more light on the phenomenon of mode change.

(7) ACKNOWLEDGMENTS

The authors are indebted to the Board of Admiralty for permission to publish the paper.

One author (G. D. S.) has been working on related theoretical problems with Dr. D. Gabor of the Imperial College Electrical Engineering Department, and would like to record his thanks for stimulating discussions on this subject.

(8) REFERENCES

(1) Willshaw, W. E., Rushforth, L., Stainsby, A. G., Latham, R., Balls, A. W., and King, A. H.: "The High-Power Pulsed Magnetron. Development and Design for Radar Applications," *Journal I.E.E.*, 1946, **93**, Part IIIA, p. 985.

(2) Gabor, D.: "Stationary Electron Swarms in Electromagnetic Fields," *Proceedings of the Royal Society*, A, 1944–5, **183**, p. 436.

(3) Hok, G.: "A Statistical Approach to the Space Charge Distribution in a Cut-Off Magnetron," *Journal of Applied Physics*, 1952, **23**, p. 983.

(4) Harris, L. A.: "Instabilities in the Smooth Anode Cylindrical Magnetron," *ibid.*, 1952, **23**, p. 562, and 1953, **24**, p. 1335.

(5) Clayton, R. J., Houldin, J. E., Lamont, H. R. L., and Willshaw, W. E.: "Radio Measurements in the Decimetre and Centimetre Wavebands," *Journal I.E.E.*, 1946, **93**, Part III, p. 97.

(6) Guénard, P., and Huber, H.: "Étude expérimentale de l'interaction par ondes de charge d'espace au sein d'un faisceau électronique se déplacant dans des champs électrique et magnétique croisés," *Annales de Radio-électricité*, 1952, **7**, p. 252.

(7) Posthumus, K.: "Oscillations in a Split Anode Magnetron," *Wireless Engineer*, 1935, **12**, p. 126.

(8) Jepsen, R. L., and Muller, M. W.: "Enchanced Emission from Magnetron Cathodes," *Journal of Applied Physics*, 1951, **22**, p. 1196.

621.396.11.029.6

Paper No. 1782 R
Jan. 1955

REVIEW OF LONG-DISTANCE RADIO-WAVE PROPAGATION ABOVE 30 Mc/s

By W. J. BRAY, M.Sc.(Eng.), Member, H. G. HOPKINS, B.Sc., Ph.D., Member, F. A. KITCHEN, B.Sc., Associate Member, and J. A. SAXTON, D.Sc., Ph.D., Member.

(*The paper was received* 27*th August*, 1954.)

SUMMARY

The purpose of the paper is to summarize the present state of knowledge concerning the factors affecting long-distance radio-wave propagation at frequencies above 30 Mc/s. Attention is drawn to various ionospheric and tropospheric propagation mechanisms and their broad characteristics. The account of these special processes is preceded by Sections dealing with selected aspects of "normal" propagation. A comprehensive Bibliography is included.

(1) INTRODUCTION

In this paper, which covers the frequency band from 30 Mc/s to 100 Gc/s,* a general account is given of the factors affecting long-distance propagation. Although a great deal of research remains to be done, it is thought that a brief review of these salient propagation features is useful at the present time, when more and more services are operating above 30 Mc/s. The paper describes some of the factors which have to be taken into account when considering long-range reception, and—perhaps more important—when considering the likely incidence of interference between well-separated stations operating on the same or neighbouring frequencies.

The account is divided into two main sections dealing respectively with transmission under "normal" conditions and transmission when special propagation mechanisms are operative. This division, although convenient for the present purpose, is artificial in that some of the features mentioned under the "normal" section (e.g. the effects of obstacles) are of importance whatever the propagation mechanism; and, conversely, the future may show that certain of the mechanisms here described as "special" are, in fact, of common and universal occurrence.

Throughout the paper the terms very high frequency (v.h.f.), ultra-high frequency (u.h.f.) and super-high frequency (s.h.f.) refer respectively to the bands 30–300 Mc/s, 300 Mc/s–3 Gc/s and 3–30 Gc/s.

(2) NORMAL PROPAGATION

(2.1) Influence of Standard Atmospheric Refraction

The refractive index of the atmosphere, which is about 1·0003 at the earth's surface, depends on the pressure, temperature and water-vapour content of the air. When the lower atmosphere is well mixed, the refractive index decreases uniformly with height by about 10^{-6} per hundred feet; these are referred to as "standard" conditions. Thus, under standard conditions of refraction there is a slight downward bending of radio-wave trajectories and the radio-horizon distance is greater than the geometrical one. The effect is equivalent to the hypothetical case of rectilinear propagation through a homogeneous atmosphere above an earth of radius 4/3 times the true physical radius. The radio-horizon distance, *D*, between two terminal aerial systems is thus given approximately by

$$D = \sqrt{(2h_1)} + \sqrt{(2h_2)} \text{ statute miles} \quad . \quad . \quad . \quad (1)$$

where h_1 and h_2 are the respective aerial heights in feet. If D is measured in kilometres and h_1 and h_2 in metres, the numerical factor is 4·14 instead of $\sqrt{2}$. The distance given by the equation is about 15% greater than the geometrical-horizon distance.

(2.2) Diffraction Around the Earth's Surface, with Standard Atmospheric Refraction

The attenuation rate with the standard atmosphere for both horizontal and vertical polarization and in the absence of atmospheric attenuation at super-high frequencies (see Section 2.5), at distances well in the diffraction region beyond the radio-horizon distance, is given by

$$\left.\begin{aligned} a &= 0\cdot152 f^{1/3} \text{ decibels per statute mile} \\ &= 0\cdot095 f^{1/3} \text{ decibels per kilometre} \end{aligned}\right\} \quad . \quad . \quad (2)$$

where f is the operating frequency in megacycles per second. Eqn. (2) applies to diffraction around a smooth earth. Thus, diffractive attenuation in the band concerned increases from about 0·3 dB/km at 30 Mc/s to 1·4 dB/km at 3 Gc/s. Such high rates of attenuation tend to restrict service ranges to approximately that of the radio horizon between the terminal aerials, particularly as the frequency is increased.

Eqn. (2) is strictly applicable only at points which are beyond the radio horizon by amounts which vary with the heights of the terminals and the operating frequency. At centimetre wavelengths the equation may, for practical purposes, be taken to apply from the horizon distance itself, but at longer wavelengths the horizon distance must be considerably exceeded. For example, at 30 Mc/s with transmitter and receiver both at a height of 200 m it must be increased by 32 km, and with both at a height of 2 000 m by 80 km; at 300 Mc/s, with both terminals at a height of 50 m it must be exceeded by 16 km, and with both at 500 m by 40 km. If the transmitter or receiver height is very high or very low compared with the other, these excess ranges increase. Thus, at 30 Mc/s, when h_1 is 200 m and h_2 is 10 km the range excess is 160 km; when h_1 is 200 m and h_2 is 0 it is 100 km: these figures will obviously be the same if transmitter and receiver heights are interchanged. At distances closer to the radio horizon than those indicated by the above examples, the attenuation rate will be less[1] than that given by eqn. (2).

In circumstances where a degraded quality of reception is acceptable, use can be made of the signals beyond the service range of the transmitter. Eqn. (2) shows that well in the diffraction region the field strength is reduced by 10 dB for an increase in distance ranging from about 35 km at 30 Mc/s to 7 km at 3 Gc/s. If other conditions are kept constant, these distances therefore indicate the order of range extension beyond the service range of the transmitter to be expected for a degradation in signal/noise ratio of 10 dB. It will be appreciated that these figures are intended only as a rough guide and there is an implicit assumption that the service range of the transmitter itself extends somewhat beyond the horizon.

---

* 1 gigacycle per second (Gc/s) = 1 000 Mc/s.

Written contributions on papers published without being read at meetings are invited for consideration with a view to publication.

This is an "integrating" paper. Members are invited to submit papers in this category, giving the full perspective of the developments leading to the present practice in a particular part of one of the branches of electrical science.

The paper is an official communication from the Radio Research Station, Department of Scientific and Industrial Research.

Mr. Bray is in the Post Office Engineering Department.

Mr. Kitchen is in the Royal Naval Scientific Service.

Dr. Hopkins and Dr. Saxton are at the D.S.I.R. Radio Research Station.

The increasing diffractive attenuation beyond the horizon with increase of frequency very largely offsets the effects of the increase in aerial gain which becomes possible with increasing frequency. Consider a given aperture having a gain of 10dB at 300Mc/s; its use at this frequency instead of an aerial with 0dB gain will, for points well beyond the horizon and for a given signal/noise ratio, increase the range of reception by 16km. Although an aperture of the same size has a gain of 30dB at 3Gc/s, the increase of reception range (over an aerial with 0dB gain) is limited to 22km by the frequency law indicated in eqn. (2).

The above examples serve as illustrations of the application and limitations of eqn. (2) which, it must be remembered, is true only for smooth-earth propagation under standard conditions of atmospheric refraction at distances well beyond the horizon.

(2.3) Effects of Surface Roughness

The discussion given above has been based on an idealized earth, i.e. one which is perfectly smooth. For transmission over the sea at the low-frequency end of the band under review this condition is most closely satisfied, but under other circumstances the structure and nature of the surface plays a large part in determining the fields experienced in practice.

The roughness of the surface reduces its reflection coefficient, so that the deep field minima at certain points within the horizon to be expected under smooth-earth conditions are less likely to occur. Observations show, in addition, that the general field strengths within the horizon are lower than would be expected on smooth-earth theory (see Section 2.4). Beyond the horizon, in the diffraction region, the roughness of the land or sea surface between transmitter and receiver upsets the ideal relation between the various terms in the expression from which the resultant field is computed. It is found that fields in this region are greater than would be expected for smooth-earth propagation and it is possible, though by no means certain, that the roughness of the terrain is a significant contributory factor.

U.H.F. measurements for standard conditions of atmospheric refraction show rather similar basic characteristics for both overland and oversea transmission. On the higher frequencies in the band, diffuse reflection takes place from most types of land or sea surface, specular reflection occurring only at almost zero grazing angle. Accordingly, there is often an appreciable reduction in the strength of the reflected wave (especially for horizontally-polarized radiation), and it has been suggested that this may contribute to an enhancement of field strength at points beyond the horizon to values greatly in excess of those calculated on the smooth-earth basis. For propagation over the sea the effective reflection coefficient varies over a wide range with the state of the sea surface, so that field strengths in the diffraction region beyond the horizon are not characteristically stable, particularly at the higher frequencies. There is similarly a general tendency toward decreased stability in transmission characteristics in overland transmission as the degree of obstruction of the propagation path is increased.

The roughness of the surface also affects the structure of the refractive-index profile of the air above it, and—particularly overland—upsets the concept of horizontal stratification. In the present state of knowledge it is not possible to say more than that field strengths computed on this basis are most likely to be reliable over the sea well removed from coastlines, or over very flat, uniform desert country.

(2.4) Effects of Large Obstacles

We are concerned with two broad categories in this treatment: (*a*) where the effect of an individual obstacle is predominant, and (*b*) where the collective effect of many obstacles is of importance. The second grouping merges into surface roughness, as discussed in the preceding Section.

For single obstacles of simple geometrical form, diffraction theory may be applied. Fresnel treatment, implying no distinction between horizontally- and vertically-polarized radiation, can be applied when the terminal points are well removed from the obstacle. Thus, a hill[2,3] may be regarded as a sharp edge at which the direct and ground-reflected waves of the incident radiation are diffracted. The field at the receiver can similarly be resolved into direct and ground-reflected rays emanating from the edge, and these waves will give rise to the usual interference pattern on and above the earth's surface. In exceptional cases where the geometry of the transmission path is just right, very strong signals can be obtained in the shadow of the ridge.[4] In general, these special conditions will not be met, but the possibility should not be overlooked that the edge-diffraction loss at a high mountain range intervening in a long transmission path may be smaller than that arising from continuous diffraction round the curved surface of the earth over a smooth path of similar length. The effects of a tree or mast may be treated in the same way,[5,6] using the Fresnel method.

When the terminal points are close to the obstacle, a more rigid approach must be made. For example, McPetrie and Ford[7] investigated the attenuation caused by bare cylindrical ridges of radii from 375m to 3·3km over a frequency range of about 30Mc/s to 3Gc/s. For horizontal polarization the measured attenuation in the shadow of the ridge agreed with the exponential law of Domb and Pryce[1] for a cylinder of the appropriate radius. For wavelengths longer than 50cm it was found that vertically-polarized radiation gave a better signal in the shadow of the ridge, whereas horizontally-polarized radiation gave a better signal outside the shadow. For wavelengths shorter than 50cm there was no observable difference between the signals obtained with vertical and horizontal polarization.

Where the collective effect of many obstacles is the determining factor, the field-strength variations from place to place will depend upon the general character of the terrain irregularities. For example, spatial variations will be much greater when there are hills which obstruct the transmission path than when there is a clear path over an undulating surface. The variations will increase with increasing frequency. Where many obstacles are involved, it is impossible to estimate the field-strength changes from place to place, and a statistical approach is necessary.[8–12] For example, it has been found that, using horizontal polarization within the radio horizon, over stretches of hilly country the median field-strength falls below the smooth-earth value by about 6dB/octave for frequencies above about 70Mc/s. The available data, however, are very limited, and it is not possible to draw any general conclusions except for the class of terrain for which most experimental results are available.

(2.5) Attenuation due to Oxygen, Water Vapour, Clouds, Fog and Precipitation at super-high Frequencies

At very high and ultra-high frequencies the atmosphere does not produce any attenuation by absorption (loss) occurring in its constituents; nor is appreciable attenuation caused by the scattering arising from "patchiness" in the permittivity structure of the atmosphere, which in turn is due to atmospheric turbulence. As will be seen later (Section 3.7), this last scattering mechanism may produce appreciable fields at long distances, but the scattering process diverts only a very small proportion of the energy and therefore introduces but little attenuation along any given radius from the transmitter.

At about 3Gc/s, however, attenuation due to absorption in the oxygen and water-vapour in the air, and absorption and scattering due to water drops and collections of ice crystals, begin to be significant and at higher frequencies become dominant factors in limiting the range of reception.

Oxygen[13–21] has a magnetic dipole moment which interacts with the magnetic field of the radiation and produces a resonance absorption band with a peak centred at about 60 Gc/s. From 3 to 30 Gc/s the oxygen absorption of the air increases from about 0·01 to 0·02 dB/km; it then rises sharply to the peak value of about 14 dB/km at 60 Gc/s and then falls, although there are other peak values at still higher frequencies.[16–21] These values refer to air at ground pressure. At frequencies well removed from, and lower than, 60 Gc/s the attenuation rate may be taken as proportional to the pressure and inversely as the square of the absolute temperature; thus the attenuation values quoted will be reduced at high altitudes. The molecule of water vapour has an electric dipole moment such that resonance, and therefore increased absorption, occurs at frequencies of about 22·5 and 187 Gc/s, owing to interaction with the electric field of the radiation.[15, 21–24]

The relative influence of oxygen and water-vapour absorption varies with the operating frequency and the amount of water vapour present. While oxygen is distributed uniformly along a given propagation path at a fixed height, the amount of water vapour present, and hence the absorption of energy due to it, may vary. The absorption due to the atmospheric gases at a given frequency can be calculated if the pressure, temperature and water-vapour content of the air are known.[13, 22] There is good agreement[15–20] between the theoretical and measured values of oxygen absorption, and similarly in the case of water vapour for the absorption band centred about 22·5 Gc/s. Some discrepancy exists between the theoretical and practical values of water-vapour absorption for the band centred about 187 Gc/s and for other bands at still higher frequencies, all of which make some contribution to the absorption in the millimetre and centimetre wavebands.

Taking various temperatures and absolute humidities of the atmosphere to be representative of typical and maximum absorption conditions in polar, temperate and tropical-equatorial regions, the associated atmospheric attenuations of radiation of various frequencies in the s.h.f. band for a total atmospheric pressure of 76 cm (1 013 mb) are given in Table 1.[25, 26]

Table 1

ATTENUATION CAUSED BY ATMOSPHERIC GASES

Frequency	Attenuation					
	Polar region		Temperate region		Tropical-equatorial region	
	Typical (0° C, W.V. 4 gm/m^3)	Maximum (5° C, W.V. 7 gm/m^3)	Typical (15° C, W.V. 10 gm/m^3)	Maximum (20° C, W.V. 17 gm/m^3)	Typical (25° C, W.V. 18 gm/m^3)	Maximum (30° C, W.V. 31 gm/m^3)
Gc/s	dB/km	dB/km	dB/km	dB/km	dB/km	dB/km
3	0·007	0·007	0·007	0·007	0·007	0·008
9·5	0·01	0·01	0·015	0·02	0·02	0·03
13·35	0·015	0·02	0·025	0·035	0·04	0·06
25	0·09	0·15	0·2	0·35	0·4	0·6

Theoretical prediction of the magnitude of the attenuation due to fog and fine-droplet clouds is in good agreement with the observed results. Owing to the small size of water droplets constituting fogs, the attenuation (in decibels per kilometre) at a given temperature and frequency in the s.h.f. band is proportional to the total mass of water per unit volume of air. To a good degree of approximation, the attenuation of radio waves may be related to the visual range in a fog, as shown in Table 2 for a temperature of 0° C.[25]

Table 2

ATTENUATION CAUSED BY FOG, EXCLUDING CONTRIBUTION DUE TO ASSOCIATED WATER VAPOUR

Frequency	Attenuation at various visual ranges in fog					
	30 m	60 m	90 m	150 m	225 m	300 m
Gc/s	dB/km	dB/km	dB/km	dB/km	dB/km	dB/km
3	0·02	0·008	0·004	0·002	0·001	0·001
9·5	0·2	0·08	0·04	0·02	0·01	0·007
13·35	0·4	0·15	0·08	0·04	0·02	0·015
25	1·2	0·45	0·25	0·12	0·07	0·045

Fogs are expected to give rather less attenuation at higher temperatures; and to obtain the attenuation in fogs of similar visual ranges in temperate (about 15° C) and tropical-equatorial regions (about 25°–30° C), the above values should be multiplied by 0·6 and 0·4 respectively. In Tables 2 and 3 the attenuations shown are for the liquid water alone; oxygen and water-vapour attenuations are additional.

The attenuation due to rain in the s.h.f. band increases with frequency and with increasing rates of precipitation, in accordance with the theoretical predictions. Since the drop-size distribution varies with the rate of precipitation, it is not possible to use a simple conversion factor to allow for variations in temperature. In Table 3, typical attenuation values are given for conditions in polar, temperate and tropical-equatorial regions.[25]

In this country a precipitation rate of 5 mm/h for one hour is likely to be exceeded on some ten occasions per year, and a rate of 50 mm/h will probably occur only for a few minutes once per year. Furthermore, it has been estimated that precipitation rates exceeding 25 mm/h are unlikely to occur over an area more than four miles in diameter, so that high attenuation will probably be confined to a small proportion of the propagation path.

In almost all circumstances likely to occur in practice, at frequencies below 30 Gc/s the attenuation produced by ice particles in the atmosphere, whether in the form of hail, dry snow or ice-crystal clouds, is very much less than that caused by rain at the same mass-rate of precipitation. For raindrops and hailstones of similar diameter (equal to 0·3 cm) the ratio between the attenuation due to hail and that due to rain at the same mass-rate of precipitation (both attenuations being expressed in decibels) is estimated to be about 0·1 at 3 and 30 Gc/s, reaching a minimum value of 0·01 in the region of 10 Gc/s.

The average cloud, containing about 0·15 gm/m^3 of liquid water, is expected to give an absorption of about 0·6 dB/km at 100 Gc/s, and a hundredth of this at 10 Gc/s. Since clouds do not, in general, exist along a large proportion of a propagation path, absorption is negligible even if the clouds are at a low temperature. Similarly, the attenuation due to ice-crystal clouds is negligible.

The deleterious effects of precipitation in relation to radar operation in the s.h.f. band are discussed in Reference 25.

(3) SPECIAL PROPAGATION MECHANISMS

In Section 2 the main propagational factors relevant to reception under normal conditions have been briefly discussed. In the present Section an outline is given of special propagation mechanisms which under certain circumstances are conducive to long-distance reception.

It is convenient to focus attention on two main regions above the earth's surface: the ionosphere extending from somewhat below 100 km to a few hundred kilometres above the surface,

Table 3

Attenuation in Rain, excluding contribution due to Associated Water Vapour

Frequency	Attenuation in various rates of rainfall																				
	0·5mm/h			1·0mm/h			2·5mm/h			5·0mm/h			12·5mm/h			25·0mm/h			50·0mm/h		
	Polar	Temperate	Tropical	Polar	Temperate	Tropical	Polar	Temperate	Tropical	Polar	Temperate	Tropical	Polar	Temperate	Tropical	Polar	Temperate	Tropical	Polar	Temperate	Tropical
Gc/s	dB/km	dB/km	dB/km	dB/km	dB/km	dB/km	dB/km	dB/km	dB/km	dB/km	dB/km	dB/km	dB/km	dB/km	dB/km	dB/km	dB/km	dB/km	dB/km	dB/km	dB/km
3	0·0004	0·0002	0·0001	0·0007	0·0004	0·0003	0·0015	0·001	0·0007	0·003	0·002	0·0013	0·007	0·004	0·003	0·015	0·008	0·006	0·03	0·02	0·01
9·5	0·005	0·004	0·003	0·01	0·01	0·009	0·025	0·03	0·03	0·06	0·07	0·08	0·15	0·23	0·22	0·35	0·55	0·55	0·8	1·3	1·2
13·35	0·015	0·015	0·015	0·03	0·03	0·03	0·08	0·09	0·07	0·18	0·25	0·2	0·5	0·65	0·6	1·0	1·4	1·3	2·2	2·9	2·7
25	0·05	0·04	0·04	0·09	0·1	0·1	0·25	0·3	0·25	0·55	0·65	0·6	1·5	1·8	1·7	3·1	3·7	3·6	6·5	7·5	7·4

and the troposphere ranging from ground level up to about 10 or 20km. In the former region it is the presence of free electrons which accounts for the refracting, reflecting and scattering properties, while in the troposphere these properties result from the non-uniform refractive-index structure of the air caused by spatial variation of pressure, temperature and humidity.[27]

Ionospheric effects will be discussed in the first four sub-sections and tropospheric mechanisms in the subsequent three. Thus the order of the sub-sections corresponds roughly to the frequency order, effects most important in the v.h.f. band being considered first.

(3.1) Propagation by way of the F2 Layer

The F2 layer—the highest in the ionosphere, with a height ranging from 250 to 400km—plays an important part in enabling world-wide communication to be carried out at high frequencies. When the maximum density of ionization within the layer is sufficiently high, waves at the low-frequency end of the v.h.f. band also may be reflected and received at large distances. As with h.f. transmission, a minimum distance (skip distance) over which reception is possible is set by the frequency of operation, the maximum density of ionization within, and the height of, the layer. The higher the frequency, the lower the maximum ionization density, and the higher the layer, the larger is the skip distance. Routine observations of the height of the layer and its maximum density are made at a number of observatories throughout the world. From these measurements at vertical incidence the reflecting properties of the layer are estimated for the very oblique angles of incidence (small grazing angle) necessary for long-distance communication. It is usual to express the maximum usable frequency (m.u.f.) at which communication between any two points can occur via the F2 layer at a given time, as the product of an m.u.f. factor and the critical frequency at the point of reflection; the critical frequency is the frequency at which penetration first occurs at vertical incidence as the frequency is increased. The m.u.f. factor strictly relates to single-hop conditions, but means are available to extend its application to multi-hop circuits. The curvature of the earth imposes a maximum range for single-hop transmission of about 4 000km. Multi-hop transmission is possible within this range and also to greater distances, but generally becomes less likely as the frequency increases, since, as the order of reflection increases, the incidence of the ray at the layer becomes steeper and waves of a given frequency can thus penetrate more readily; furthermore, multi-hop transmission to great distances can occur only when there is sufficiently dense ionization within the layer over wide geographical areas.

In the F2 layer, ionization conditions suitable for v.h.f. propagation occur most frequently at the maximum of the sunspot cycle. It is considered that circumstances will arise only extremely rarely when F2 propagation is possible above about 70Mc/s, and this limit frequency will be lower as the range is closed from about 4 000km.

(3.2) Propagation by way of Sporadic-E Ionization

Long-range propagation of waves at the lower end of the v.h.f. band can occur as the result of abnormally dense ionization appearing in a thin stratum of the E region at about 100km above the earth's surface. This is termed sporadic-E ionization (Es) since it generally exhibits highly irregular variation with time and position. Its importance lies in the fact that it occurs low down in the ionosphere and that high ionization-density gradients are involved. Consequently m.u.f. factors (ranging up to 5·6) for Es propagation are much higher than the corresponding F2 values for a given length of circuit. On the other hand, the critical frequency is generally, but not always, lower than the F2 value, reflection may be only partial and the layer is seldom dense over wide areas; multi-hop transmission is accordingly most unlikely and the maximum range of transmission cannot much exceed 2 000 km because of the limitation set by the curved surface of the earth. As with propagation by way of the F2 layer, a skip distance exists within which reception will not generally be possible, but in contrast with F2 propagation, points within the skip may occasionally be reached by very-high-power transmitters; this is because weak partial reflections can occur in the Es case, whereas for F2 transmission all of the energy incident more steeply than the angle corresponding to the skip distance penetrates the layer and is not returned to the earth. Except under rare circumstances and when very high power is used, the minimum Es skip distance is expected to be about 600 km at 30 Mc/s, rising to 1 500 km at 54 Mc/s and 2 000 km at 60 Mc/s. Furthermore, at a given distance (greater than the skip distance) the reflection coefficient of the layer decreases with increase of frequency. Thus, for a given transmitter power the range of distances from which reception via the Es layer is possible falls steadily and becomes very small for frequencies about 60 Mc/s. These examples are representative of conditions in Western Europe. As an example of the exceptional conditions which can arise when very high power is used at the transmitter, it may be noted that for about half an hour in one day during the summer of 1953 the Kirk o'Shotts television signals on 53 Mc/s were intercepted via the Es region at a range of 550 km. It may be remarked that the intensity of ionization in the Es layer in this part of the world increases and decreases in sympathy with normal E-ionization; consequently long-distance interception via the Es layer is more likely by day than by night and in local summer months. On the other hand, in high latitudes the reverse is true and limiting frequencies are higher than those mentioned above.

(3.3) Propagation by way of Scattering from the E Region

Scattering from the E region[28] produces weak, rapidly fluctuating, but substantially continuous signals at metre wavelengths from a few hundred kilometres up to a maximum distance of about 2 000 km set by the curvature of the earth. The importance of this propagation mechanism arises from its availability on occasions when other of the mechanisms under consideration are absent. In the foregoing Sections it has been possible to use the picture of an ionospheric layer behaving as a reflecting surface and having special properties of reflection in respect of frequency and conditions of incidence. But in practice the ionization density is not homogeneous and a scattered component of radiation exists as well as the reflected signal. When the operating frequency exceeds the m.u.f., the energy which would have been reflected to the receiving point penetrates the E layer, but the weak scattered component remains to provide a possible means of communication. The scattering centres are regions of more-than-average electron density in a state of more or less random relative motion,[29] believed to be distributed over a thin layer about 100 km above the earth's surface. Although the energy impinging on this stratum is mainly scattered forward (transmitted) along the direction of incidence, a small proportion of it is deflected and may be intercepted at a distant receiving point provided that the portion of the scattering region illuminated by the transmitter is "visible" from the receiving position. Since the received signal is the resultant of a very large number of contributions of varying amplitude and phase, a received carrier appears to be modulated by a band of noise a few cycles wide, and pulse-modulated signals are lengthened by perhaps 10 microsec owing to the delay imposed on the more widely deviated scattered components.

For given transmitting and receiving aerials directed at the scattering region over the mid-point of the path, theory suggests[28, 30] that the received signal-power depends on the following factors:

(*a*) The thickness of the scattering layer: the thicker it is, the more centres are illuminated and the higher the field strength.

(*b*) The magnitude of the mean-square fractional variation in electron density: the more intense the spatial fluctuations within the scattering layer, the greater the signal.

(*c*) The scale of turbulence, roughly corresponding to the average size of the blobs of high ionization-density.

(*d*) The angle at the scattering centre between the direction of incidence (direction from transmitter) and the direction of scatter (receiver direction): as this angle of deviation increases from zero (forward scattering) the signal from a given scattering centre falls off very rapidly.

(*e*) The frequency of the transmitter: the received signal power falls off very rapidly with increase of frequency through the v.h.f. band.

(*f*) The distance of the transmitter: the received signal power would not be expected to vary rapidly with distance, since the reduction in signal strength due to increasing distance is offset by the necessarily associated reduction in the angles of deviation at the various scattering centres. Thus, from a maximum distance of about 2 000 km (determined by the height of the scattering region and the curvature of the earth) down to a few hundred kilometres, no very great change in received signal power would be expected.

(3.4) Propagation by way of Aurorae

The intense ionization produced at the base of auroral streamers contributes to the properties of the Es layer at high latitudes (Section 3.2). It is possible in addition to obtain reflections at the low-frequency end of the v.h.f. band from the streamers themselves[31–34] which may give rise to long-distance propagation. The streamers may be considered very crudely as long corrugated conducting sheets having lateral dimensions large compared with the wavelength and following the general direction of the earth's magnetic field. In the high latitudes where aurorae are of most frequent occurrence the earth's field has a strong vertical component, so that the general direction of the sheets will usually be less than 25° to the vertical. Auroral displays occur most frequently at a height of about 100 km; there is a sharp lower boundary to the distribution, but no corresponding well-marked upper limit, auroral effects being observable on occasion at heights of several hundred kilometres.

For a large received signal-power the transmitter and receiver should be beamed at the streamers, and reflection will occur from those parts of the streamer surfaces which are suitably disposed relative to the transmitter and receiver directions. The reflection points must be "visible" from both transmitter and receiver; that is, with an arbitrary maximum height of 400 km, each terminal must be within 2 000 km of the streamer and the maximum distance between transmitter and receiver cannot exceed 4 000 km. Much smaller transmitter-receiver separations would, however, be expected to be more favourable, because the above ranges are based on a height of 400 km, whereas 100 km is more representative; and, on the basis of the crude model, the multiplicity of intervening sheets between transmitter and receiver at extreme transmitter-receiver separations might be expected to preclude reflection from any suitably disposed sheet situated along the line joining transmitter to receiver, except perhaps when both terminals are only just outside the auroral belt. In practice, the smaller transmitter-receiver separations appear more favourable, the maximum separation reported being 1 200 km at 50 Mc/s;[33] there is no evidence of skip effects. It should be noted in conclusion that recent experimental work on 35 and 74 Mc/s[35] strongly suggests that, at least in some circumstances, long-distance propagation may result, not from reflection from the aurorae themselves, but by scattering from the ground on the far side of the auroral display. On this basis, propagation is by way of an enhanced Es layer. Clearly more experimental data are required before the auroral propagation process is fully understood.

Experience shows that the signals propagated by the auroral mechanism are badly mutilated by extremely rapid fading, with components of the order of tens or even hundreds of cycles per second.[32] Telephony may be practicable at 30 Mc/s and occasionally at 50 Mc/s; at 144 Mc/s, however, it is unintelligible, although telegraphy at hand speeds is readable.

In the northern hemisphere[36] the contour of most frequent occurrence of aurorae overhead is a circle centred in the region of the geomagnetic pole and having a radius of about 20–25°. Thus it extends as far south as Labrador (60° N), but in longitude 70° E it reaches only the northernmost part of Novaya Zemlya (77° N). The corresponding contour in the southern hemisphere is believed, on the basis of rather meagre information, to be about 20° in radius;[36, 37] it extends as far north as 55° S in the general longitude of Tasmania, but in the sector 30° E through 0° to 140° W over Antarctica it is probably at latitudes higher than 80° S. Observations made in the northern hemisphere show considerable variation from place to place in the time of night at which aurorae occur most frequently. It is to be expected that radio reception statistics will reflect this variability. The great majority of reported occurrences of v.h.f. communication via aurorae refer to the night hours, but cases occurring before sunset and after sunrise are known.[32]

(3.5) Propagation by way of Elevated Inversions

In the preceding four sub-sections we have been concerned with propagation mechanisms connected with an ionized medium. It has been seen that these effects are most important at the lower frequencies of the band under consideration. Attention will now be directed to long-distance mechanisms arising from the properties of the un-ionized atmosphere in the troposphere.

The first of these propagation mechanisms is reflection at an extensive and fairly sharp discontinuity in the gradient of refractive index with height, occurring well above the earth's surface. It is sometimes found that, although the lapse of refractive index with height near the ground may be close to the standard value (Section 2.1), at heights ranging between about 90 m and, say, 3 km considerable and—from the radio-transmission point of view—significant increases over the normal lapse rate may occur. Both the magnitude and extension in height of these departures may vary over a wide range. Such conditions may be produced, for example, by the occurrence of a subsidence inversion, including that which sometimes occurs immediately below a cold frontal surface. These regions of greater-than-normal rate of decrease in refractive index with height are associated with a less-than-normal rate of decrease in temperature with height, often amounting to an inversion; but the factor which mainly determines the refractive-index variation in the region is the change in humidity lapse-rate which is very frequently associated with the abnormal temperature variation.

The effect of these "inversion layers," as we may conveniently call them, may in many cases be determined to a close approximation by a simple application of ray theory, on the assumption that the layers produce partial and occasionally total reflection of the radiation. In practice, inversion layers often extend over a height of many wavelengths, and in these circumstances the reflection coefficient is appreciably less than it would be if the integrated departure of the refractive index from standard over the layer were concentrated in an abrupt discontinuity. Even so, the reflection coefficient of practically occurring inversion layers is sufficient to be of importance in the long-range transmission of radiation in the frequency band 30–300 Mc/s and perhaps higher. The efficiency of reflection falls off with increasing frequency, because the changes in refractive index with height, expressed in terms of the change per wavelength, become more gradual as the frequency rises.

It is clear that, in the absence of other propagation mechanisms, the transmitter and receiver locations will have to be "visible" from the region of reflection at the layer. From this simple viewpoint and taking a maximum height for the layer of 3 km, the maximum distance over which the mechanism can operate is about 500 km for a ground-based transmitter and receiver and 250 km for a layer of 750 m; but to these ranges must be added the horizon ranges corresponding to the heights of the terminals [eqn. (1)], and at the lower frequencies, diffraction may increase the effective ranges considerably.

Multi-hop signals are unlikely to occur for ground-based terminals, even if the inversion layer is sufficiently extensive, because the reflection coefficient is usually very small. Estimates of the reflection coefficient show that, when the grazing angle is sufficiently small (a very small fraction of a degree), total reflection may occur (cf. total internal reflection in the optical case), but at steeper angles, reflection is only partial. For a given profile of refractive index in the region of the inversion, the critical glancing angle decreases as the frequency is increased. For fixed transmitter and receiver positions, a given inversion layer might hypothetically provide strong signals at, say, 40 Mc/s but only a weak reflection at 100 Mc/s. If the height of the layer gradually increased, the strength of the higher-frequency signal would decrease more rapidly than the other. It should be noted, however, that the condition of total reflection is unlikely to occur in practice for ground-based terminals, as the glancing angle required is extremely small. It might be achieved in the case of a low inversion or by an inversion with a sloping surface. The latter possibility cannot be ruled out, but uncertainty about the appropriate meteorological details and lack of sufficient experimental radio data prevent more being said at present.

In conclusion it may be remarked that the increase of signal strength produced by an elevated inversion for ground-ground transmission may introduce some reduction when the reception or transmission point is above the inversion. In special circumstances this may lead to trouble in communication between aircraft and the ground,[40] namely when the inversion is intense and when the aircraft, although above the inversion, is at an extremely low angle of elevation.

(3.6) Propagation under conditions of Super-Refraction near the Earth's Surface

When the atmosphere is well mixed (such as usually occurs, for example, in unsettled cyclonic conditions and poor weather) the temperature, water-vapour content and, of course, pressure decrease steadily with increasing height. These conditions produce a steadily decreasing value of refractive index as height above the ground increases and result in a ray-curvature downwards of about one-quarter that of the earth. This state of affairs is taken as standard and is equivalent to rectilinear propagation in a medium of constant refractive index above an earth having an effective radius of four-thirds its geometrical value (see Section 2.1). When the meteorological conditions immediately above the earth's surface are such that the rate of decrease in refractive index with height is greater than standard, super-refraction exists.[41–43]

Conditions of super-refraction near the earth's surface may be broadly divided into two classes. In the first (which is the more important in the frequency band 30–300 Mc/s) there is a general increase in the negative gradient of refractive index with respect to height as compared with standard conditions: this leads to a greater curvature of the radio-ray paths, and to a greater effective flattening of the earth (i.e. greater than corresponds to 4/3 times the geometrical radius) with the rectilinear-propagation concept. Secondly, the decrease of refractive index with height may become so great, over a sufficient but limited range of height, that a radio duct is formed. When this happens, radio energy from a transmitter situated within the duct, at angles near to the horizontal, becomes effectively trapped if the frequency is sufficiently high, and is propagated with a relatively high intensity to much greater distances than would be possible without the duct. The lower the frequency the thicker the duct required to give appreciable trapping of radio energy, and even at 300 Mc/s a duct must extend to a height of a few hundred feet before the effect becomes of great importance, although partial trapping occurs with somewhat smaller ducts. Although ducts having thicknesses of a few hundred feet seldom occur in the British Isles, they can be formed in certain areas, e.g. over sea in tropical regions, in the Mediterranean and in the Middle East.

Booker and Walkinshaw[44] have developed a theoretical treatment of radio ducts close to the earth's surface. The phenomenon may be described by a series of characteristic E or H waves, the modes of which have track widths which increase with the order of the wave. Usually the earth forms the lower edge of the track, and the upper edge of any particular mode is determined by the refractive-index profile. Beyond the horizon the diffracted field takes the form of a partially-guided wave, whose first-mode track extends up to about 90 m at 300 Mc/s and 9 m at 3 Gc/s. Energy is fed into the duct at the horizon, and is guided round the curved surface of the earth below the horizon. It is the leakage from the top of the duct that determines the degree to which the wave is guided, and it is the refractive-index lapse rate within the track of the first mode which primarily determines this leakage. Occasionally at ultra-high frequencies, leakage of the first mode is almost completely suppressed, resulting in the production of very high field-strengths in its track at long ranges. Super-refraction diverts to the region below the horizon some of the energy that would otherwise have been radiated near the horizon itself and thus lost.

As the frequency is increased above the u.h.f. band, the occurrence of ducts of thickness and intensity adequate to provide trapping of at least one mode becomes more frequent and potentially of considerable importance. On the other hand, it must be remembered that the effects of surface roughness become more pronounced as the frequency is increased, and these may either (from meteorological considerations) prevent the establishment of an extensive duct or (from radio considerations) impose an attenuation term which would not be present over an ideally smooth earth. It may be mentioned here that the semi-permanent low-level ocean duct[45, 46] up to some tens of feet high is of importance at frequencies above 2 or 3 Gc/s. The duct is the result of evaporation from the sea surface; the strata in the immediate neighbourhood of the surface have a high water-vapour content which decreases very rapidly with height.

In conclusion, it may be remarked that, of the special propagation mechanisms so far discussed, this is the first which shows, from the viewpoint of long-distance propagation, a rising characteristic with frequency. It is not unfair to add that meteorological conditions are seldom so simple that quantitative correlation and comparison are possible between theoretical expectations and practical radio results.[47–49] There is, however, no reason to doubt the validity of the broad descriptions given above.

(3.7) Propagation by way of Scattering from Atmospheric Turbulence

Until comparatively recently it was generally accepted that, in the absence of the other special mechanisms already described, the field strength beyond the horizon would fall exponentially with increasing distance [eqn. (2)] in accordance with diffraction theory. Also that, for frequencies in the upper part of the u.h.f. band and higher, the attenuation would be so great that little increase in effective range would be obtained by practicable increases in power or aerial directivity (Section 2.2). Numerous measurements over both land and sea paths have shown that, under conditions known to approximate to those of a standard well-mixed atmosphere, the attenuation rate for some distance beyond the horizon agrees quite closely with that predicted by diffraction theory. At somewhat greater ranges, however, field strengths have been observed[50–53] to be much higher than would be expected on the basis of conventional theory for a standard atmosphere. Pekeris[54] suggested that the slowly attenuated fields might be produced by a scattering process, and pointed out that the fluctuating character of the signal at long range was consistent with this supposition. This mechanism was already recognized as the cause of one type of fading commonly experienced at near-horizon ranges. The suggestion of the mechanism of scattering of radio waves in the lower atmosphere has been interpreted by several workers.[30, 55–59]

In Section 3.3 the effects of scattering centres in the ionosphere have been discussed; in this case, scattering results from patchiness in ionization, with consequent inhomogeneity of refractive index, and the effective inhomogeneity falls off with increase of frequency. However, if spatial fluctuations in refractive index of suitable linear dimensions and intensity were available in the un-ionized part of the atmosphere, owing to highly local variations in humidity, temperature or pressure, the rapid decrease in effectiveness of scatter with increase of frequency would be absent. Such a state of affairs exists in the troposphere, owing to atmospheric turbulence, and provides a mechanism for long-distance propagation.

Atmospheric turbulence results from the frictional forces generated by surface drag and windshear, and the buoyancy forces derived from heat received at the earth's surface by solar radiation. During day-time heating the two forces act together in producing the resultant turbulent fluctuations, but during the night, cooling of the earth tends to cause stability in the surface layers and thus reduces turbulence, at any rate in these layers. The observed fluctuations are primarily a result of vertical mixing processes between air at different levels having different average properties. Although practical measurements of turbulence do not extend above 3 km, it seems probable that turbulent effects at levels much above 9–12 km (except perhaps under jet-stream conditions of rather limited lateral extent or from fully developed thunder clouds in the tropics) will have only small effects on radio propagation, because at great heights the low density of the atmospheric constituents imposes a limitation on the magnitude of the refractive-index variation available. Practical measurements of the fluctuations of refractive index in the atmosphere are described in References 60 and 61.

The various theories developed start from the assumption that the lower atmosphere is permanently in a state of turbulence which produces and supports local departures in refractive index from the mean value. These local inhomogeneities generally have some horizontal uniformity. They may be regarded as a system of scattering centres which, when illuminated in the region of the horizon, produce an essentially forward-scattered beam directed well into the diffraction region. Meteorological data on the eddy size and distribution in the atmosphere are slight, and as a result the various theoretical interpretations referred to have arisen from the use of different statistical descriptions of the turbulence.

In Booker and Gordon's approach[30] the spatial scale of the small fluctuations in refractive index is described by a space-correlation function. It is shown that, in the region beyond the horizon of the transmitter, the available field consists of a component due to refraction and one due to scattering, the latter being often as strong as, or stronger than, the refraction component. At distances just beyond the horizon the refracted wave is predominant and the scattered wave is then a cause of fading. As the distance increases, the effect of scattering predominates: one result of this is an apparent broadening of the polar diagram of a high-gain receiving-aerial system.

In the treatment by Lagrone[59] the general equation for scattering developed by Booker and Gordon is extended to a volume integral equation, which gives the total scattered power density at a receiving point relative to the power radiated per unit solid angle by an isotropic source. Numerical application of the equation shows that maximum scattered power occurs from the direction of the transmitter, low down on the receiver horizon.

Staras,[58] whilst using an integral expression for the scattered power equivalent to that used by Booker and Gordon, states that it is not appropriate to use a space-correlation function of refractive-index variation as defined by these authors. Instead, he defines a time-correlation function of refractive index which permits formal evaluation of the time-average scattered power. It is also shown that, whereas a space-correlation function of refractive index is not readily evaluated by experiment, the time-correlation function is directly measurable. Finally, it is shown that, for so-called small-scale turbulence, the average scattered power does not depend appreciably on any particular model of atmospheric turbulence, whereas for large-scale turbulence the frequency and scattering-angle dependence of the scattered energy is greatly affected by the particular time-correlation function chosen.

In Megaw's approach[56, 57] the 3-dimensional fluctuations of refractive index, assumed statistically stationary, are supposed analysed into a spectrum of triply-periodic components. For a small interval of the spectrum a representative half-space-wavelength cube is considered as a scattering source, and the field is evaluated for the most favourable phasing in the three co-ordinates. It has been assumed that the appropriate space integration, with random phasing between the elements, gives the correct contribution to the scattered field. Whilst there is some doubt about the correctness of the final coefficients, the theory retains the constants specifying a finite spectrum in the expression for the scattered field, after the final integration over the spectrum. The treatment indicates that, allowing for the decrease in aerial directivity which commonly accompanies an increase in wavelength, the r.m.s. scattered field, expressed as a fraction of the free-space field, does not vary widely over the whole frequency range for which it is observable. Megaw has conducted extensive trials of the reception of the field scattered by atmospheric turbulence, on a frequency of 3 Gc/s on oversea paths up to more than 550 km. The results of the trials were in reasonable agreement with the theoretical predictions and with other observations made under quite different circumstances. From this it was concluded that scattering due to turbulence in the atmosphere would, in some degree, be present everywhere and at all times.

While the theories of scattering from atmospheric turbulence of Booker and Gordon and of Megaw start from different viewpoints, they can be shown to be formally identical. However, to explain their radio results Booker and Gordon chose a scale of turbulence about one-thousandth of that which is now known to be appropriate. As a consequence, some of their conclusions are in error.

Thus, although there is a lot to learn about the details of this propagation mechanism, it is already clear that it provides a means of transmission inherently available up to the top of the s.h.f. band. The upper range limit will be set by geometrical conditions—scattering region "visible" from both transmitter and receiver, allowing some addition for diffraction around the bulge of the earth. Its practical effects at frequencies above a few gigacycles per second, however, will be severely restricted by the attenuation factors discussed in Section 2.5.

(4) ACKNOWLEDGMENTS

The authors are indebted to Messrs. H. G. Cundall and E. Fitch for their assistance and helpful discussions.

This report is published by permission of the Admiralty, the Engineer-in-Chief of the Post Office and the Director of Radio Research, Department of Scientific and Industrial Research.

(5) BIBLIOGRAPHY

(1) DOMB, C., and PRYCE, M. H. L.: "The Calculation of Field-Strengths over a Spherical Earth," *Journal I.E.E.*, 1947, **94**, Part III, p. 325.

(2) SCHELLING, J. C., BURROWS, C. R., and FERRELL, E. B.: "Ultra-Short-Wave Propagation," *Proceedings of the Institute of Radio Engineers*, 1933, **21**, p. 427.

(3) BULLINGTON, K.: "Radio-Propagation above 30 Mc/s," *ibid.*, 1947, **35**, p. 1123.

(4) DICKSON, F. H., EGLI, J. J., HERBSTREIT, J. W., and WICKIZER, G. S.: "Large Reductions of V.H.F. Transmission Loss and Fading by the Presence of a Mountain Obstacle in Beyond-Line-of-Sight Paths," *ibid.*, 1953, **41**, p. 967.

(5) MEGAW, E. S. C.: "Some Effects of Obstacles on the Propagation of Very Short Radio Waves," *Journal I.E.E.*, 1948, **95**, Part III, p. 97.

(6) MCPETRIE, J. S., and FORD, L. H.: "Some Experiments on the Propagation of Radiation of 9·2 cm Wavelength, especially on the Effect of Obstacles," *ibid.*, 1946, **93**, Part IIIA, p. 531.

(7) MCPETRIE, J. S., and FORD, L. H.: "An Experimental Investigation on the Propagation of Radio Waves over Bare Ridges in the Wavelength Range 10 cm to 10 m," *ibid.*, p. 527.

(8) BROWN, G. H., EPSTEIN, J., and PETERSON, D. W.: "Comparative Propagation Measurements," *RCA Review*, 1948, **9**, p. 177.

(9) ALLEN, E. W.: "U.H.F. Propagation Characteristics," *Electronics*, 1949, **22**, p. 86.

(10) BULLINGTON, K.: "Radio Propagation Variations at V.H.F. and U.H.F.," *Proceedings of the Institute of Radio Engineers*, 1950, **38**, p. 27.

(11) SAXTON, J. A.: "Basic Ground-Wave Propagation Characteristics in the Frequency Band 50–800 Mc/s," *Proceedings I.E.E.*, Paper No. 1602R, January 1954 (**101**, Part III, p. 211).

(12) SAXTON, J. A., and HARDEN, B. N.: "Ground-Wave Field-Strength Surveys at 100 and 600 Mc/s," *Proceedings I.E.E.*, Paper No. 1601R, January 1954 (**101**, Part III, p. 215).

(13) VAN VLECK, J. H.: "The Absorption of Microwaves by Oxygen," *Physical Review*, 1947, **71**, p. 413.

(14) BERINGER, R.: "The Absorption of One-Half Centimetre Electro-Magnetic Waves in Oxygen," *ibid.*, 1946, **70**, p. 53.

(15) LAMONT, H. R. L.: "Atmospheric Absorption of Millimetre Waves," *Proceedings of the Physical Society*, 1948, **61**, p. 562.

(16) STRANDBERG, M. W. P., MENG, C. Y., and INGERSOLL, J. C.: "Microwave Absorption Spectrum of Oxygen," *Physical Review*, 1949, **75**, p. 1524.

(17) BERINGER, R., and CASTLE, J. G.: "Microwave Magnetic Resonance Absorption in Oxygen," *ibid.*, p. 1963.

(18) BUCKHALTER, J. H., ANDERSON, R. S., SMITH, W. V., and GORDY, W.: "Fine Structure of Microwave Absorption Spectrum of Oxygen," *ibid.*, 1950, **79**, p. 651.

(19) ANDERSON, R. S., SMITH, W. V., and GORDY, W.: "Line Breadths of Fine Structure of Microwave Spectrum of Oxygen," *ibid.*, 1951, **82**, p. 264.

(20) ANDERSON, R. S., JOHNSON, C. M., and GORDY, W.: "Resonant Absorption of Oxygen at 2·5mm Wavelength," *ibid.*, 1951, **83**, p. 1061.

(21) SHOSTAK, A.: "Atmospheric Absorption Chart," *Electronics*, 1952, **25**, p. 134.

(22) VAN VLECK, J. H.: "The Absorption of Microwaves by Uncondensed Water Vapour," *Physical Review*, 1947, **71**, p. 425.

(23) BECKER, R. E., and AUTLER, S. H.: "Water Vapour Absorption of Electro-Magnetic Radiation in the Centimetre Wavelength Range," *ibid.*, 1946, **70**, p. 300.

(24) SAXTON, J. A.: "The Dielectric Properties of Water Vapour at Very High Frequencies," Report of a conference on "Meteorological Factors in Radio-Wave Propagation" (Physical Society, London, 1947, p. 215).

(25) SAXTON, J. A., and HOPKINS, H. G.: "Some Adverse Influences of Meteorological Factors on Marine Navigational Radar," *Proceedings I.E.E.*, Paper No. 1060R, January 1951 (**98**, Part III, p. 26).

(26) RYDE, J. W.: "The Attenuation and Radar Echoes produced at Centimetre Wavelengths by Various Meteorological Phenomena," Report of a conference on "Meteorological Factors in Radio-Wave Propagation" (Physical Society, London, 1947, p. 169).

(27) SAXTON, J. A.: "The Propagation of Metre Radio Waves beyond the Normal Horizon: Part 1—Some Theoretical Considerations with Particular Reference to Propagation over Land," *Proceedings I.E.E.*, Paper No. 1112R, March 1951 (**98**, Part III, p. 360).

(28) BAILEY, D. K., *et al.*: "A New Kind of Radio-Propagation at Very High Frequencies Observable over Long Distances," *Physical Review*, 1952, **86**, p. 141.

(29) RATCLIFFE, J. A.: "Diffraction from the Ionosphere and the Fading of Radio Waves," *Nature*, 1948, **162**, p. 9.

(30) BOOKER, H. G., and GORDON, W. E.: "A Theory of Radio Scattering in the Troposphere," *Proceedings of the Institute of Radio Engineers*, 1950, **38**, p. 401.

(31) MOORE, R. K.: "Aurora and Magnetic Storms," *Q.S.T.*, 1951, **35**, No. 6, p. 14.

(32) MOORE, R. K.: "A V.H.F. Propagation Phenomenon associated with Aurorae," *Journal of Geophysical Research*, 1951, **56**, p. 97.

(33) GERSON, N. C.: "Radio Observations of the Aurora on Nov. 19th, 1949," *Nature*, 1951, **167**, p. 804.

(34) CHAPMAN, S.: "The Geometry of Radio Echoes from Aurorae," *Journal of Atmospheric and Terrestrial Physics*, 1953, **3**, p. 1.

(35) HARANG, L., and LANDMARK, B.: "Radio Echoes observed during Aurorae and Geomagnetic Storms using 35 and 74Mc/s Waves Simultaneously," *ibid.*, 1954, **4**, p. 322.

(36) HARANG, L.: "The Aurorae," International Astrophysics Series, Vol. 1 (Chapman and Hall, London, 1951).

(37) GARTLEIN, C. W., and MOORE, R. K.: "Southern Extent of Aurora Borealis in North America," *Journal of Geophysical Research*, 1951, **56**, p. 85.

(38) SAXTON, J. A., LUSCOMBE, G. W., and BAZZARD, G. H.: "The Propagation of Metre Radio Waves beyond the Normal Horizon: Part 2—Experimental Investigations at Frequencies of 90 and 45Mc/s," *Proceedings I.E.E.*, Paper No. 1114R, March 1951 (**98**, Part III, p. 370).

(39) SMYTH, J. B., and TROLESE, L. G.: "Propagation of Radio Waves in the Lower Troposphere," *Proceedings of the Institute of Radio Engineers*, 1947, **35**, p. 1198.

(40) PRICE, W. L.: "Radio Shadow Effects produced in the Atmosphere by Inversions," *Proceedings of the Physical Society*, 1948, **61**, p. 59.

(41) BOOKER, H. G.: "Elements of Radio-Meteorology: How Weather and Climate cause Unorthodox Radar Vision beyond the Geometrical Horizon," *Journal I.E.E.*, 1946, **93**, Part IIIA, p. 69.

(42) MEGAW, E. C. S.: "Experimental Studies of the Propagation of Very Short Radio Waves," *ibid.*, p. 79.

(43) MCPETRIE, J. S., STARNECKI, B. J., JAROWSKI, H., and SICINSKI, L.: "Oversea Propagation on Wavelengths of 3 and 9cm," *Proceedings of the Institute of Radio Engineers*, 1949, **37**, p. 243.

(44) BOOKER, H. G., and WALKINSHAW, W.: "The Mode Theory of Tropospheric Refraction and Its Relation to Waveguides and Diffraction," Report of a conference on "Meteorological Factors in Radio-Wave Propagation" (*loc. cit.*, p. 80).

(45) MCPETRIE, J. S., and STARNECKI, B. J.: "Low-Level Atmospheric Ducts," *Nature*, 1948, **162**, p. 818.

(46) ATTWOOD, S. S. (Editor): "Radio-Wave Propagation Experiments," Summary, Technical Report of the Committee on Propagation, N.D.R.C., 1946, **2**, p. 33.

(47) MILNES, B., and UNWIN, R. S.: "A Radio-Meteorological Investigation in the South Island of New Zealand," *Proceedings of the Physical Society*, B, 1950, **63**, p. 595.

(48) MACFARLANE, G. G.: "The Application of a Variational Method to the Calculation of Radio-Wave Propagation Curves for an Arbitrary Refractive Index Profile in the Atmosphere," *ibid.*, 1948, **61**, p. 48.

(49) HAY, H. G., and UNWIN, R. S.: "Extension of the Mode Theory of Tropospheric Refraction to cover Variations in the Refractive Index Profile along a Transmission Path," *ibid.*, 1952, **65**, p. 981.

(50) DAY, J. P., and TROLESE, L. G.: "Propagation of Short Radio Waves over Desert Terrain," *Proceedings of the Institute of Radio Engineers*, 1950, **38**, p. 165.

(51) KATZIN, M., BAUCHMANN, R. W., and BINNIAN, W.: "Three and 9cm Propagation in Low-Level Ocean Ducts," *ibid.*, 1947, **35**, p. 891.

(52) NORTON, K. A.: "Advances in Electronics" (Academic Press, Inc., New York, 1948), Vol. 1, p. 381.

(53) STRAITON, A. W., METCALF, D. F., and TOLBERT, C. W.: "A Study of Tropospheric Scattering of Radio Waves," *Proceedings of the Institute of Radio Engineers*, 1951, **39**, p. 643.

(54) PEKERIS, C. L.: "Wave Theoretical Interpretation of Propagation of 10 centimetre and 3 centimetre Waves in Low-Level Ocean Ducts," *ibid.*, 1947, **35**, p. 453.

(55) RICE, S. O.: "Statistical Fluctuations in Radio Field-Strength far beyond the Horizon," *ibid.*, 1953, **41**, p. 274.

(56) MEGAW, E. C. S.: "Scattering of Electro-Magnetic Waves by Atmospheric Turbulence," *Nature*, 1950, **166**, p. 1100.

(57) MEGAW, E. C. S.: "Waves and Fluctuations," *Proceedings I.E.E.*, 1953, **100**, Part III, p. 1.

(58) STARAS, H.: "Scattering of Electro-Magnetic Energy in a Randomly Inhomogeneous Atmosphere," *Journal of Applied Physics*, 1952, **23**, p. 1152.

(59) LAGRONE, A. H.: "Volume Integration of Scattered Radio Waves," *Proceedings of the Institute of Radio Engineers*, 1952, **40**, p. 54.

(60) CRAIN, C. M., and GERHARDT, J. M.: "Measurements of the Parameters involved in the Theory of Radio Scattering in the Troposphere," *ibid.*, p. 50.

(61) CRAIN, C. M., DEAM, A. P., and GERHARDT, J. R.: "Measurement of Tropospheric Index-of-Refraction Fluctuations and Profiles," *ibid.*, 1953, **41**, p. 284.

621.317.382.029.6:621.3.016.3

Paper No. 1743 R
Jan. 1955

A VERY-WIDE-BAND DUMMY LOAD FOR MEASURING POWER AT VERY-HIGH AND ULTRA-HIGH FREQUENCIES

By W. HERSCH, B.Sc.(Eng.), Associate Member.

(*The paper was first received* 16*th March, and in revised form* 23*rd July,* 1954.)

SUMMARY

The principles underlying the design of a wide-band dummy load are discussed. An example is described which has a purely resistive input impedance of 75 ohms at all frequencies above 100 Mc/s and is capable of measuring powers up to 1 kW with an accuracy of ±(2½% of measured power + 3 watts).

It consists of a short coaxial line, filled with carbon tetrachloride, the inner conductor being made up from short sections of thin gold-film resistors.

(1) INTRODUCTION

Numerous methods are available for measuring power at high frequencies, the choice depending upon the frequency as well as the order of magnitude of the power to be measured.

Low powers are readily determined by observing the change in resistance of a temperature-sensitive element owing to the power which it absorbs from an electromagnetic field; thermistors, bolometers and enthrakometers are typical devices of this kind.

Other effects, such as the mechanical force exerted on a diaphragm due to radiation pressure, or the relative brightness of bulbs, have been used successfully at appropriate frequencies and power levels.

The measurement of power becomes progressively more difficult as the band over which the device has to be operated is extended, and also as soon as the power level exceeds a few tens of watts.

In the first case the problem is largely one of maintaining a good match to the source, whilst in the second, adequate heat dissipation becomes of prime importance.

(2) THEORETICAL CONSIDERATIONS

(2.1) General

The requirement for a wide-band resistive dummy load as a laboratory tool arose out of a B.B.C. contract for three transmitters to cover bands III, IV and V respectively (174–960 Mc/s).

The direct measurement of large powers at video and ultra-high frequencies has been accomplished successfully at spot frequencies, using calorimetric methods whereby the power absorbed by a dummy load is determined in terms of the temperature rise of a known quantity of liquid.

There are two basic methods whereby r.f. energy can be converted into heat. We can either use the resistive losses that occur in the conductors or the loss occurring in the dielectric with which the device is filled. It is also possible, of course, to use a combination of the two.

A calorimetric method of determining power requires that a cooling liquid be circulated through the load, and water has been the obvious choice in the past. In fact it has been so exclusively used that the terms "dummy load" and "water load" have become synonymous.

Water becomes increasingly "lossy" at higher frequencies, and existing water loads rely entirely on the dielectric loss in the water to generate the heat.

Unfortunately the dielectric loss of water is by no means constant, and changes with frequency in such a manner that it is inherently unsuitable for very-wide-band applications, largely because of the difficulty of preserving a good match to the power source.

As a matter of fact, a design of a wide-band dummy load based on dielectric loss is fundamentally wrong since there are very few dielectrics that have a loss angle inversely proportional to frequency, and those that have introduce negligible losses, even at the higher frequencies, and are therefore useless.

On the other hand, resistive losses are comparatively small, and unless they are artificially increased, a dummy load capable of dissipating even modest powers has to be long, with a consequent loss of accuracy owing to heat losses by convection, quite apart from being unwieldy.

(2.2) Special Precautions demanded by Wide-Band Devices

The design of a very-wide-band dummy load to be described was based on the conclusions drawn from the foregoing Section. Briefly, it took the form of a short coaxial line, filled with carbon tetrachloride, the inner conductor consisting of a very thin metallic film on a glass rod, along which nearly all the loss takes place.

The design centred around the well-known expression for the input impedance of an infinitely long transmission line:

$$Z_0 = \sqrt{\left(\frac{R + j\omega L}{G + j\omega C}\right)} \quad . \quad . \quad . \quad . \quad . \quad (1)$$

Considering frequencies for which $\omega L > R$ and $\omega C > G$, the condition for resistive input impedance requiring that

$$\arc \tan \frac{\omega L}{R} = \arc \tan \frac{\omega C}{G} \quad . \quad . \quad . \quad . \quad (2)$$

cannot be maintained at low frequencies since R is purposely made large and G is practically zero. The input impedance will therefore have a reactive component.

This can be made very small, but the choice of R is a compromise between the permissible mismatch at the lowest frequency at which it is intended to use the load, which calls for R to be small, and the desire to make the load physically short, which requires R to be large.

The input impedance tends to become purely resistive with increase in frequency, and in order to prevent any variation of the impedance it is essential that none of the four constants L, C, G and R be dependent upon frequency, and that the ratios $\omega L/R$ and $\omega C/G$ should linearly increase with frequency.

The inductance and capacitance per unit length are determined by the physical constants as well as the permeability of the metal and the permittivity of the insulation.

Variation of permeability with frequency can easily be avoided by excluding ferromagnetic materials from the construction. The dielectric constant of most insulating materials hardly changes with frequency—with the exception of water, where a steady decrease above 5 000 Mc/s takes place. As pointed out earlier, the dielectric loss which determines G is for most

---

Written contributions on papers published without being read at meetings are invited for consideration with a view to publication.

Mr. Hersch is at the Mullard Research Laboratories.

dielectrics a function of frequency, and all dielectrics for which $\tan\delta$ is large and does not decrease linearly with frequency are to be avoided in the construction of wide-band devices because of the effect that the change of the ratio C/G has on the phase angle of the input impedance.

The range of liquid dielectrics which fulfil the above condition or which have negligible losses is consequently limited, and includes benzene, carbon tetrachloride, tetrachlorethylene, aromatic hydrocarbons, mixtures of paraffins and naphthenes, and silicones.

The choice mainly depended on the dielectric constant. In a coaxial type of dummy load solid dielectric supports must be included, and in order to avoid discontinuities, the dielectric constant of the liquid has to be as nearly as possible equal to that of the solid supports. Furthermore, the permittivity of both dielectrics, and in particular that of the liquid, must not change with temperature in the range over which the device is expected to work.

The resistance R is subject to change with frequency owing to skin effect, but by making the inner conductor in the form of a glass rod on to which a very thin film of gold—a few Ångström units thick—is deposited, the effect is virtually eliminated. This construction has two further advantages.

In order to keep the dimensions of the dummy load to reasonable proportions, it was found that the resistance had to be of the order of 1–10 ohms/cm.

Also, since nearly all the loss takes place on the inner conductor, its surface had to be made as large as possible to assist the cooling. Both requirements were conveniently met by a thin metallic film.

(3) RESISTANCE CHANGE REQUIRED FOR A LINE HAVING UNIFORM POWER DISSIPATION

For transmission lines which have a predominantly resistive characteristic impedance, the attenuation per unit length is given by

$$\text{Loss in decibels} = 8\cdot686\left(\frac{R}{2Z_0} + \frac{GZ_0}{2}\right) \quad . \quad . \quad (3)$$

Also by definition

$$\text{Loss in decibels} = 10\log_{10}\frac{P_{in}}{P_{in} - P_{lost}} \quad . \quad . \quad (4)$$

If P_{lost} is also taken to apply to a unit length of line the two expressions can be equated with one another.

In the line under consideration, the dielectric loss G is negligibly small compared with R, so that

$$10\log_{10}\frac{P_{in}}{P_{in} - P_{lost}} = 8\cdot686\frac{R}{2Z_0} \quad . \quad . \quad . \quad (5)$$

The permissible power loss per unit length depends upon how rapidly heat can be removed from the resistive conductor. In the case of a carbon-tetrachloride-cooled gold-film resistor, values of 30 watts/cm^2 can be achieved with adequate cooling.

Eqn. (5) is general and applies to any section of the line. However, P_{in} is only known at the input, and hence the value of R which can be calculated from it is the resistance per unit length at the input end.

It can easily be seen that for subsequent sections R obeys the law

$$R = R_{in}\varepsilon^{\frac{x}{L-x}} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (6)$$

where L and x are defined by Fig. 1.

If the line is correctly terminated by a disc having a radial resistance equal to the characteristic impedance, the dummy load can be shortened by an amount depending upon the power dissipation of the terminating disc.

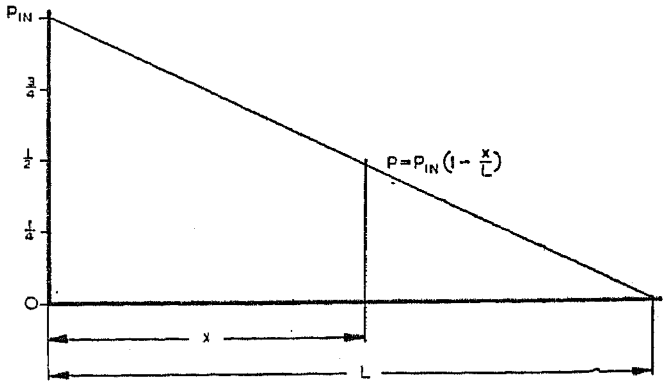


Fig. 1.—Line with uniform power dissipation.

(4) DESCRIPTION OF DUMMY LOAD

(4.1) Design

A 75-ohm dummy load capable of dissipating 1 kW and giving a standing-wave ratio of better than 1·05 at frequencies above 100 Mc/s was constructed, having an overall length of less than 4 ft.

The liquid and solid dielectrics were carbon tetrachloride and polythene respectively, the dielectric constants of which differ by less than 4%. (2·17 compared with 2·26.)

The use of carbon tetrachloride required a closed cooling system, and special precautions against corrosion had to be taken. A pair of large-capacity petrol pumps of the diaphragm type are capable of producing a flow rate of up to 100 litres/hour. Coarse control is achieved by using either one or both pumps, and fine control by means of a hand-controlled valve. A flow meter consisting of a float in a vertical tapered glass tube enables the flow to be measured with an accuracy of ±2%. A water-cooled heat exchanger maintains the inlet temperature of the carbon tetrachloride constant. Separate thermometers for inlet and outlet are provided. Fig. 2 shows the schematic of the complete cooling circuit.

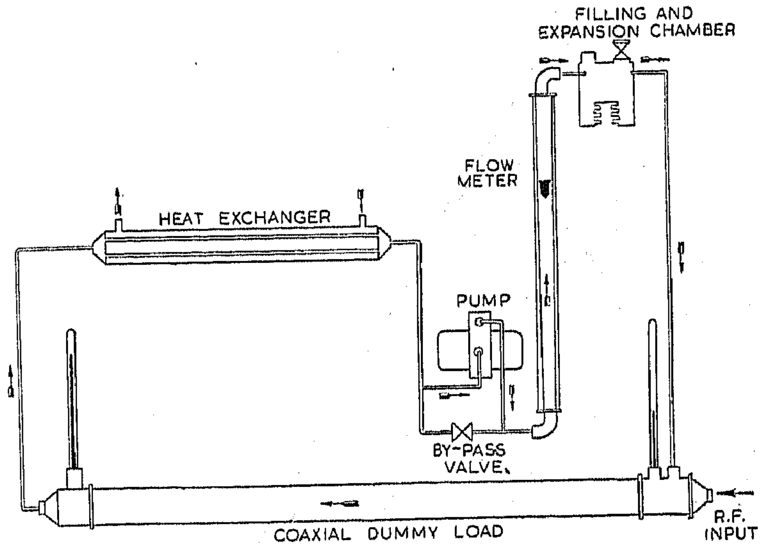


Fig. 2.—Schematic of closed cooling circuit.

The resistance of the inner conductor is graded in order to achieve uniform power dissipation, and for manufacturing reasons it was made up from short sections, each about 10 cm long.

Fig. 3 shows how the theoretical grading was approximated by eight long and four short sections.

The gold-film resistors are only 0·5 cm in diameter, and although they can handle 45 watts per centimetre of length when cooled by a liquid, the design of the dummy load was based on 6 watts/cm, the power dissipation at which the carbon tetrachloride in immediate contact with the resistor just begins to boil.

An enlarged cross-section through a gold-film resistor is shown

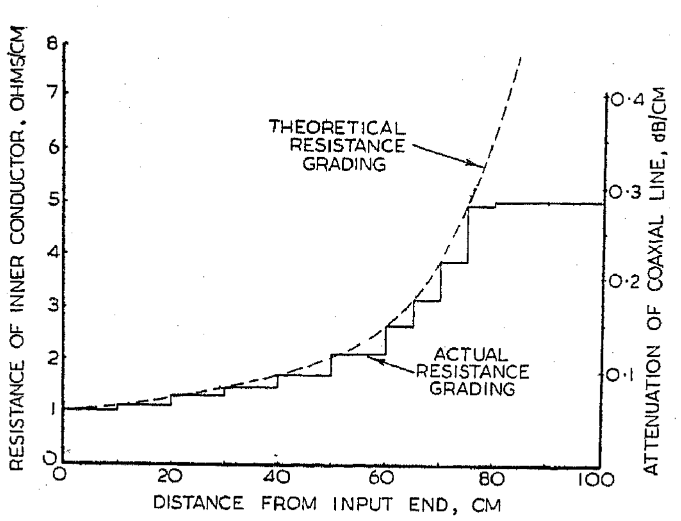


Fig. 3.—Graph showing theoretical and actual resistance grading to achieve uniform power dissipation.

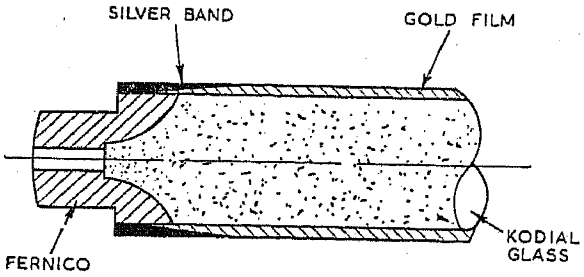


Fig. 4.—Enlarged cross-section through a gold-film resistor.

in Fig. 4. It consists of a glass rod, fused to Fernico terminations, the assembly being ground to the correct size. The gold is applied in liquid form and baked on by firing, the desired resistance being obtained by varying the thickness of the film. The gold is firmly bonded to the Fernico ends by a band of silver, also fired on. Individual sections are joined by thin silver-plated beryllium-copper sleeves which form the centre of perforated polythene spacers; thus the whole assembly is supported centrally within the outer conductor. A circular perforated disc, made of glass and Fernico and coated with platinum, with a radial resistance of 75 ohms, terminates the "far-end" of the dummy load.

(4.2) Frequency Limitations

There is no definite frequency below which the dummy load could not be used, since this depends upon the permissible departure of the input impedance from the nominal value.

However, ωL should always be greater than R in order to justify the simplifying assumption on which eqn. (3) is based.

At very-high frequencies the dummy load can support higher-order modes, the first one being a circumferential TE mode occurring when the average circumference is approximately equal to a wavelength.

For the load described in the paper, that frequency was calculated to be 3 680 Mc/s, but the fact that higher-order modes can be supported does not necessarily restrict the use of the load, for unless these modes are excited they cannot be troublesome, and by proper design, mode conversion due to discontinuities can be avoided.

(4.3) Accuracy

The overall accuracy is affected by heat losses and instrument errors. The combined effect of the latter can be reduced below 1%, although this was not done in the model described in the paper since the need for accuracy was not very great.

Heat losses are due to convection losses, changes of ambient temperature, insufficient mixing of the fluid dielectric and reflection.

The first three depend on operating conditions, and their effect was determined experimentally by replacing the graded gold-film resistors by a glass tube of the same outside diameter wound uniformly with resistance wire.

With direct current, a correction curve was obtained for the different operating conditions normally encountered.

Reflection losses were deduced from the standing-wave-ratio measurements.

(5) PERFORMANCE

The standing-wave ratio was measured by means of a solid-dielectric slotted line, having the same physical dimensions and electrical constants as the dummy load itself. At first, the s.w.r. rose to a maximum of 1·2 at 800 Mc/s, which was due to the cumulative effect of the reflections from the junctions of those resistors which were all of the same length. By making the resistors all of different length, a s.w.r. of better than 1·05 was measured at all frequencies above 100 Mc/s. No measurements were carried out above 2 500 Mc/s.

The low specific heat of carbon tetrachloride allowed a very large range of input powers to be handled and also resulted in high sensitivity. By adjusting the flow, powers from 10 watts to 1 000 watts were measured, the error being ± (2½% of measured power + 3 watts).

(6) CONCLUSIONS

Adequate heat dissipation is the main problem in connection with compact high-power dummy loads.

Wide-band characteristics can be achieved by paying careful attention to the electrical and mechanical properties of the materials employed in the construction and by avoiding electrical or physical discontinuities.

The use of thin gold-film resistors cooled by a low-loss liquid dielectric offers outstanding advantages over existing techniques, because it enables high attenuation to be combined with large surface area.

The accuracy of the dummy load described in the paper is capable of considerable improvement, since it is limited chiefly by the accuracy with which the flow rate and the temperature rise can be determined.

(7) ACKNOWLEDGMENTS

The author is greatly indebted to Mr. J. T. Balkwill, who handled all the glass work, for his skill and patience, to Messrs. Brown-Greaves and Joanes for their help and suggestions to achieve perfection of the gold-film resistors and to the management of Mullard Research Laboratories for permission to publish the paper.

(8) REFERENCES

(1) Ramo and Whinnery: "Fields and Waves in Modern Radio" (John Wiley and Sons, 1944).
(2) Terman, F. E.: "Radio Engineers' Handbook" (McGraw-Hill Book Co., Inc., 1950).
(3) Everitt, W. L.: "Communication Engineering" (McGraw-Hill Book Co., Inc., 1937).
(4) "Reference Data for Radio Engineers," Federal Telephone and Radio Corporation, 1949.
(5) Lange, N. A.: "Handbook of Chemistry" (Handbook Publications Inc., Sandusky, Ohio, 1946).
(6) Hartshorn, L., Parry, J. V. L., and Rushton, E.: "The Dielectric Losses in some Representative Insulating Materials," *Proceedings I.E.E.*, (Paper No. 1451 M, February 1953), **100**, Part IIA, p. 23.

621.396.11.029.65 : 621.373.423

Paper No. 1761 R
Jan. 1955

EXPERIMENTAL EQUIPMENT AND TECHNIQUES FOR A STUDY OF MILLIMETRE-WAVE PROPAGATION

By W. E. WILLSHAW, M.B.E., M.Sc.Tech., Member, H. R. L. LAMONT, M.A., Ph.D., Associate Member, and E. M. HICKIN, Associate Member.

(*The paper was first received* 21*st May, and in revised form* 30*th August*, 1954.)

SUMMARY

Apparatus has been developed to study the characteristics of radio-wave propagation in the region of the oxygen absorption band (wavelengths, 4–6mm). The sources of power were crystal harmonic generators and pulse and c.w. magnetrons, and the design of crystal detectors and mixers is described. The construction of waveguide components for this wavelength range is discussed, together with equipment for measurement of impedance, frequency, power and attenuation. For most of the propagation measurements, a transmitter with an output of a few microwatts was used, together with a calibrated superheterodyne receiver. Communication tests were carried out with a pulse-frequency-modulated magnetron transmitter.

Measurements of the oxygen attenuation over paths of a few kilometres have been made at a number of wavelengths in the absorption band. Measurements over sea have given values of the reflection coefficient of the sea surface. These results are summarized and discussed briefly.

(1) INTRODUCTION

To the many known complications that beset the path of radio waves through the atmosphere, another was added when it was discovered by Van Vleck[1] that the oxygen molecule shows a resonance effect which gives rise to absorption of electromagnetic radiation. Van Vleck's prediction was made on theoretical grounds, and showed that the absorption arose from the permanent-magnetic moment associated with the oxygen molecule; fortunately the effect is limited to a number of discrete wavelength regions of which the longest is in the neighbourhood of 5mm. By the time his paper appeared the effect had been demonstrated in the laboratory by Beringer,[2] who used a length of waveguide filled with pure oxygen and mixtures of oxygen and nitrogen, and measured directly the attenuation thus produced. From this work it appeared that for frequencies at the peak of the absorption curve the attenuation of a radio wave in the atmosphere would be of the order of 14dB/km.

Further work by Van Vleck[3] had shown that a similar absorption effect was caused by water vapour. This time the effect was dependent on a different mechanism—the electric dipole moment of the water molecule—and the wavelength region involved was around 1·35cm, where there is a minor maximum absorption, and in an extended series of bands covering wavelengths from 3mm downwards. The curves of Fig. 1 show the theoretical predictions for these two effects, and it is clear that between them they form an effective barrier to atmospheric propagation over any considerable distance of wavelengths in the millimetre-wave region.

At that time apparatus was available to check the water-vapour effect at 1·35cm, but almost nothing was available for measurements at millimetre wavelengths. Beringer's work was done with the aid of a 1cm oscillator followed by a crystal distorter generating the second harmonic. It was therefore decided in 1945 to institute a programme of measurements on atmospheric propagation at millimetre wavelengths in the region of the oxygen absorption, as an extension of the work on radio-wave propagation which had previously been carried out at wavelengths[4] between 10 and 3cm. The paper describes the development of transmitting, receiving and measuring equipment. With this equipment, measurements have been made of atmospheric absorption and of propagation characteristics over sea. These measurements, which were carried out between 1946 and 1949, are surveyed briefly.

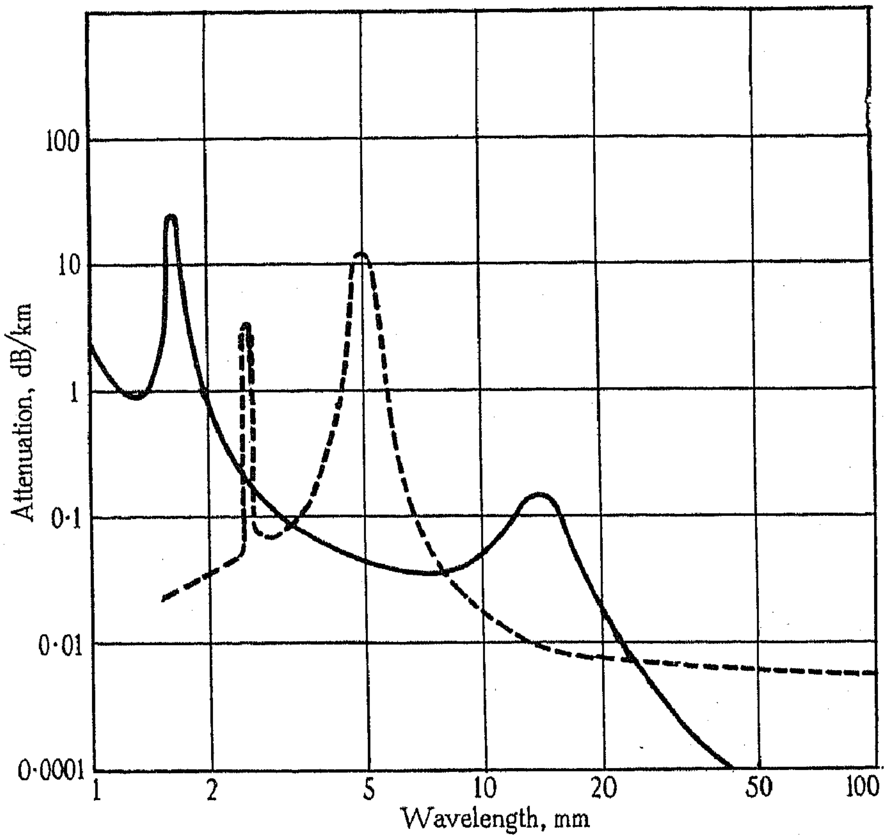


Fig. 1.—Calculated attenuation of millimetre waves due to water vapour and oxygen in the atmosphere.
—— Water vapour (7·5g/m^3).
---- Oxygen ($\Delta\nu/c$ = 0·02cm^{-1}).

(2) APPARATUS

(2.1) Signal Sources

(2.1.1) Crystal Distorters.

When this work was commenced there were no millimetre-wave oscillators, and the shortest wavelength source available was the VX302 reflex klystron, operating between 1·225 and 1·275cm and giving an output of a few milliwatts into a waveguide output circuit. A second harmonic generated from this would have a wavelength just at the skirt of the oxygen absorption band of Fig. 1. To enable measurements to be made up to the peak of the oxygen absorption, two modified versions of the VX302 were produced. These were the VX5015 and VX5016, covering respectively the wavebands 1·025–1·075 and 1·125–1·175cm.

Written contributions on papers published without being read at meetings are invited for consideration with a view to publication.

The paper is a communication from the staff of the Research Laboratories of The General Electric Co., Ltd., Wembley, England.

Dr. Lamont is now with RCA (in London).

In order to generate the second harmonic, a crystal distorter was used. Several methods of matching the fundamental power into the crystal and extracting the second harmonic were tried, the most satisfactory being illustrated in Fig. 2. The crystal cube was mounted flush with the top surface of a waveguide, whose dimensions were suitable for propagation of the second harmonic but not of the fundamental. The whisker mounting and choke system were fitted to the under surface. This distorter was mounted close to the oscillator klystron, and the length of waveguide between valve and crystal was filled with a tightly fitting length of ground fused quartz whose permittivity was sufficiently great to allow the waveguide to transmit the fundamental wavelength. The quartz was tapered at the crystal end, and the whisker passed through a hole drilled at the base of the taper. A wave trap was fitted between the valve and the crystal to prevent harmonic power from passing back into the valve. By this method, most of the fundamental power was fed into the crystal, and rectified currents of about 10mA were obtained. Comparisons were made between silicon and germanium material, and while little difference in conversion efficiency was found, crystals using germanium proved to be more stable, especially with a platinum-ruthenium whisker welded to the surface. With the valves mentioned above, second-harmonic outputs of up to about 200μW were detected, corresponding to a conversion efficiency of about 2%.

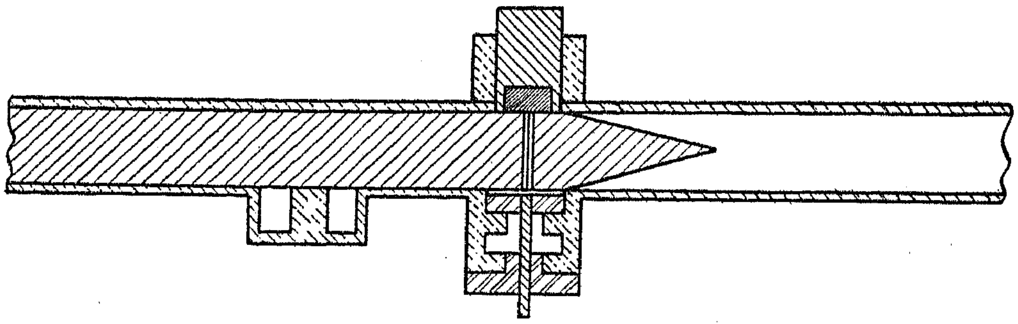


Fig. 2.—Crystal-distorter circuit using quartz-loaded waveguide.

(2.1.2) Pulse Magnetron.

Although progress could be made with such small powers, and indeed the greater part of the measurement programme was carried out with these sources, it was felt most desirable to have a fundamental oscillator in the 5–6mm range. The necessarily small size set big problems in design for both magnetrons and klystrons. It was the general opinion that this wavelength was about the lowest possible for the existing type of velocity-modulation valve, and that a magnetron would offer better chances. It was therefore decided to investigate the possibility of making a pulsed magnetron, the aim being to produce a number of valves to enable propagation tests to be carried out at spot frequencies in the oxygen absorption band. A mean output of a few watts, with a peak output of a few kilowatts, was desirable.

For wavelengths of the order of 5–6mm it was considered that any type of strapped structure would present insurmountable constructional difficulties, and that a rising-sun structure would be the most suitable. With parallel-sided slots, flexibility of design could be achieved, since wavelength changes from valve to valve could be achieved easily by variation of slot lengths, and in addition the mode spectrum of the structure could be varied considerably by altering the ratio of long to short slots.

After a series of experiments on 18-segment slotted rising-sun structures, it was decided to change to a 28-segment block, since this would enable both the anode-hole and cathode diameter to be increased appreciably. Initially these 28-segment blocks had a circuit ratio of 1·6 : 1 (i.e. ratio of long to short slot length), but as a result of further experiment it was concluded that the mode separation was too small, and blocks were constructed with a larger circuit ratio (2·5 : 1). This value was chosen as giving maximum mode separation consistent with mechanical design.

The slot lengths were varied from valve to valve in order to vary the wavelength, but the circuit ratio was maintained near 2·5 : 1.

Subsequent development consisted primarily in improving the cathode and output circuit, and improving the external mechanical construction of the valve, now known as type VX3046.

The valves operated in a magnetic field of 12 000–14 000 oersteds (supplied by a pair of C-shaped magnets attached to the valve pole pieces), with a pulse length of 0·1 microsec and a repetition rate of 10kc/s. The maximum input was 70 watts (70kW peak) with a peak voltage and current of 14kV, 5amp, the input being limited by cathode dissipation. For the same pulse length but at the lower repetition rate of 1kc/s they could run at 100–200kW peak input. The efficiency with 70 watts mean input varied from 2% to 10% with a mean value around 4·5%. Of the valves tested, the life at 70 watts input was only a few hours, failure being due in all cases to loss of cathode emission.

Valves have been made for wavelengths in the range from 5·39mm to 6·25mm with substantially the same performance, dimensions being scaled down for those in the range 5·39–5·74mm. An output window having glass approximately half a wavelength thick performed satisfactorily at each wavelength without design change.

No mode change was observed on any of the valves, at least up to a current of 15amp, except when the cathode emission was on the point of failing.

It was evident that one of the reasons for the somewhat low efficiency indicated above was the relative looseness of the valve-to-output coupling, and some increase would be required in order to produce a more efficient source. Cathode-dissipation problems require further consideration to enable the valve to handle more input and to increase the valve life.

(2.1.3) Continuous-Wave Magnetron.

While the pulse magnetron was suitable as a transmitting valve, it was not useful as an oscillator for measurement work, so after experience with the pulse magnetron it was decided to attempt to design a c.w. magnetron with an output of a few milliwatts in the 5–6mm wavelength region. The obvious difficulty lay in the small size of both anode and cathode required, but as a very small output power would suffice, an investigation was started into the possibility of operating at the lowest possible anode voltage, i.e. in the region where the efficiency approaches zero. This would allow a maximum size of anode and cathode for a given operating voltage. It was thought that a smaller ratio of cathode-to-anode diameter than usual would be necessary. With these considerations in mind it was decided to make use of the 28-segment rising-sun anode structure already developed for pulse operation. The minimum voltage at which this could be expected to oscillate was about 3·6kV, with a magnetic field dependent on the cathode diameter but of the order of 3 500 oersteds.

The first experiments were made with a model of the 28-segment block scaled to 3·2cm, the anode diameter being 16mm. Cathodes of different diameters were tried, giving cathode-to-anode ratios between 0·03 and 0·41, and an optimum value of about 0·19 was found. This gave an output efficiency of 5% with an anode voltage[5] of the order of 4·5kV.

A 6mm valve was then made with a tungsten spiral cathode of 0·6mm outside diameter accurately centred by supports at both ends. This gave a power output of 600mW at a wavelength of 6·06mm when the valve was operated with a 50c/s anode voltage applied through a 10 000-ohm series resistor. After further experiments the block was reduced in size to bring the wavelength down to about 5·6mm, and the cathode-to-anode ratio altered to a final value of 0·26. Under best conditions a maximum power of 1 watt has been obtained with a 1 : 1 square-wave anode-voltage supply, the efficiency being about 2% at a voltage of 4·3kV with a magnetic field of 3 600 oersteds.

The life obtained with tungsten spiral cathode was short, the maximum being 15h, and experiments were made with cathodes having thoria coatings. A design of this type gave a life of 60h. A further difficulty which aggravated the cathode-life problem was the occurrence of very high bombardment at a current below that at which normal oscillation started. The ill effect of this could be avoided by square-wave modulation, and c.w. operation was then possible by starting the valve with relatively short pulses and slowly increasing the "on" time to infinity. A further improvement should be gained by increasing the anode and cathode length, and this will also reduce the back bombardment.

No measurements of noise have yet been made on this magnetron with regard to its eventual use as a local oscillator, but it is interesting to note that measurements on valves operated in the "minimum-voltage" regime at 3cm wavelength show that the noise generated is no greater than in a klystron oscillator.

(2.1.4) Velocity-Modulation Oscillators.

When this project commenced it was generally believed that velocity-modulation oscillators could not be made satisfactorily for wavelengths as short as 5mm. However in 1949 a few samples became available. One valve gave 2mW output at 5·75mm and ran for 80h.

Had sources of this nature been available earlier, the work described in the paper would undoubtedly have been accelerated.

(2.2) Crystal Rectifiers

In making crystals for use as mixers or detectors it was clearly impossible to scale down the normal form of capsule crystal used at centimetre wavelengths to such a size that it would fit in a waveguide whose broad dimension was 4·5mm. A way out of the difficulty was found by designing a crystal[6] to plug into a cavity resonator which was tuned to a higher mode than the fundamental and was thus of larger size.

The crystal unit itself formed part of a flat cylindrical E_{020} resonator, as shown in Fig. 3, the top surface of which was formed by a disc at whose centre was mounted the crystal cube. The lower pressing, into which the whisker assembly was soldered, formed part of the lower surface of the resonator. The two parts were joined by sealing to a length of ceramic tube. In the E_{020} mode the radial variation of electric field is as indicated in Fig. 3(*a*), the radial length being equivalent to three-quarters of a wavelength. The diameter of the ceramic tube was therefore chosen so that it lay at the zero of the electric field, i.e. at the quarter-wavelength point, in order to reduce losses to a minimum.

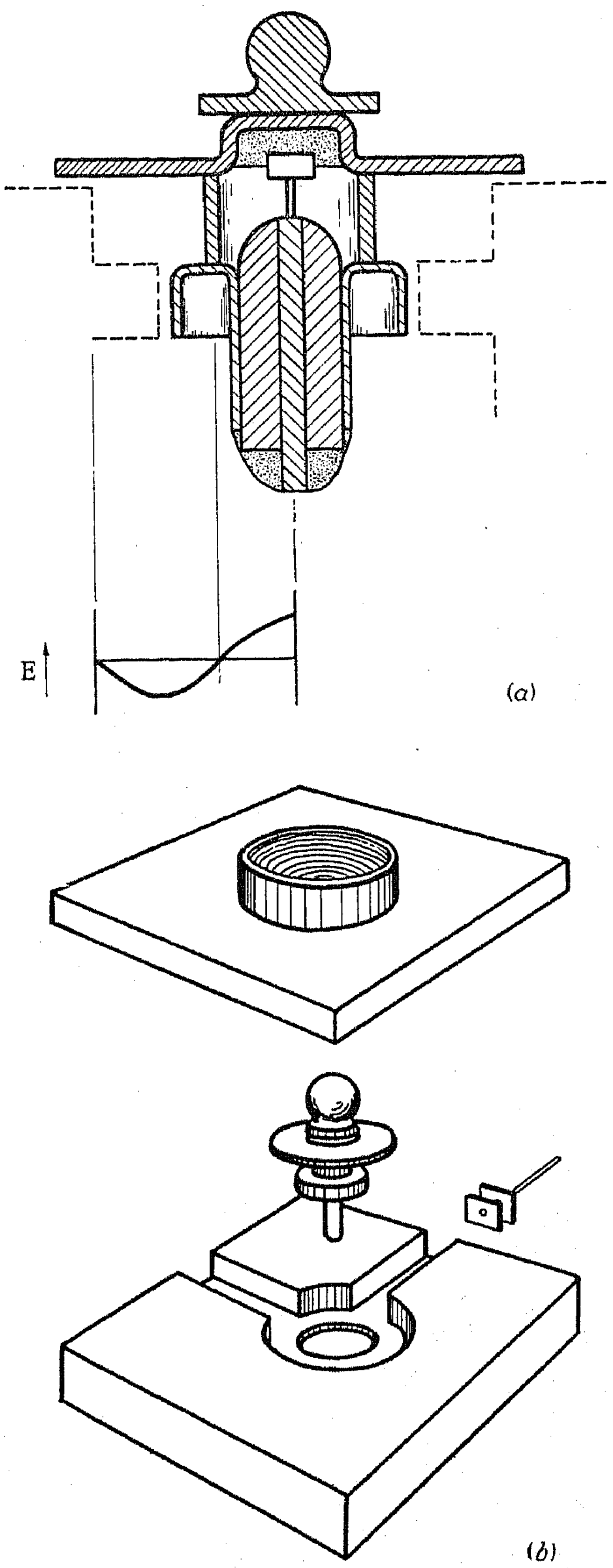


Fig. 3.—Crystal rectifier and circuit.
(*a*) Cross-section showing potential distribution.
(*b*) Exploded view of crystal and circuit.

The lower pressing was bent over so as to form, with the wall of the resonator, a capacitor across which direct or modulation voltages were established. The diameters were chosen in this case so that the gap lay at the half-wave point, where the wall current is zero, thus causing minimum disturbance to the field.

Fig. 3(*b*) shows the resonator cavity, with the crystal removed, in the form used for a detector. Its depth was the same as the small dimension of the waveguide. A waveguide coupling channel was milled into the block and a side stub with tuning plunger was also provided. An r.f. rejection filter was constructed in the d.c. output connection.

Where the crystal was used as a mixer, a second waveguide-coupling channel was milled in the block for the local-oscillator signal.

Another type of detector used was based on the VX3089 crystal (designed for 8 mm use). Here the crystal was mounted in one wall of the guide, the whisker approaching from a door-knob protrusion on the opposite wall. The whisker mounting was isolated to enable a coaxial lead to be connected.

The VX3089, although nominally of 8% bandwidth at about 8 mm, was also found to be satisfactory if the back plunger was readjusted.

(2.3) Waveguides and Small Waveguide Components

In the early work on this programme a rectangular waveguide of internal dimensions 4·5 × 2·5 mm was used. In order to be in line with other standard guides this was later changed to one of internal dimensions 0·180 × 0·090 in, and this standard was accepted by the Radio Components Standardization Committee* as waveguide No. WG24.

The coupling designed for use with this guide is shown in Fig. 4. It is of the symmetrical, screwed-ring type now adopted as standard at longer wavelengths, and uses two broached discs, 1, accurately aligned by a keyed ring, 3, and clamped by two screwed rings, 2.

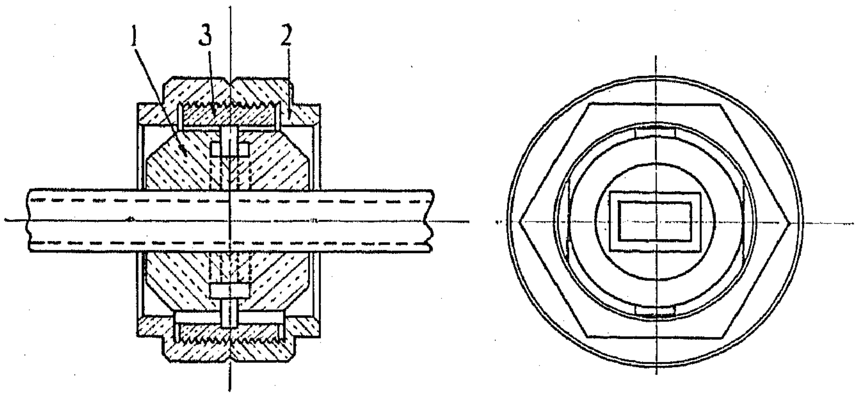


Fig. 4.—Coupling for millimetre waveguide (WG24).

Various materials were used as terminating loads—wood, conducting rubber and cellulose acetate loaded with iron powder. The latter gave standing-wave ratios better than 0·95 with a 1 in taper on a 3 in length of material.

Rotating joints were also made in which two guides were coupled through a divided section of circular guide with a choke ring at the joint. No detailed standing-wave-ratio measurements were made, but the input impedance varied little with rotation of the joint.

Tuning pistons have been of the usual type having two quarter-wave plungers with quarter-wave separation, the leading plunger being of the non-contact type.

Variable reactances have also been constructed, consisting of two quartz stubs with $\frac{3}{8}$ wavelength separation, which can be moved along the guide by one screw adjustment and altered in penetration by another.

The design of the above components does not differ in principle from those used at longer wavelengths, but the dimensions are such that the highest quality of workmanship is required to maintain a sufficiently small dimensional tolerance.

(3) R.F. MEASUREMENT TECHNIQUES

As part of the programme, apparatus had to be made for measurements on transmitters, receivers and components in the 5 mm wavelength region. These are described in the following Subsections.

(3.1) Impedance

For consistent results the accuracy of movement required in a standing-wave detector of the moving-probe type is very high at all wavelengths. At millimetre wavelengths the requirements are particularly severe, and consequently, as has already been mentioned, the highest possible accuracy of workmanship is called for. It was considered that the difficult problem of ensuring that the probe moved in its slot with the minimum of transverse movement, and of producing an extremely narrow and uniform slot, would best be met by machining the slotted section in the form of an arc of a circle of large radius, the probe carriage being carried on an arm pivoted at the centre of this circle. This construction is indicated in the views of Fig. 5. The slotted guide is formed from the two parts, 1 and 2, both machined in the same operation, and subsequently pinned together. The radius to the centre of the guide is 5 in. The probe, shown in detail in Fig. 5(*c*), is carried on the arm, 3, which makes contact with the slotted guide only on the surface, 4, against which it is held by spring pressure. Choke slots, 5, are included in the lower surface of the arm. The probe is shielded by an arc-shaped ridge, 6, partially filling the slot, extending for some distance on either side of the probe, and tapered to a point at each end. The probe passes into a waveguide, 7, energy from the probe being transferred to this guide through a dome-shaped transformer, 8. To the remote end of this guide a crystal detector is attached.

There is little information on the maximum allowable size of slots in standing-wave detectors if detectable radiation is to be avoided. Common practice may be taken to be given by the figures of Hirst and Hogg,[7] who quote a slot width of 0·063 in as satisfactory at 3 cm wavelength. Scaling this down to 6 mm wavelength gives a slot width of 0·013 in. It was not possible to attain this figure without omitting the probe shield, but the satisfactory figure of 0·025 in was found to be possible. The probe was a glass-sheathed 0·005 in wire sliding in a 0·010 in hole, the ridge, through which the probe passes, having a width of 0·020 in. The probe was adjustable by means of a differential screw, 9, and the carriage moved by the micrometer, 10.

With an instrument of this type, standing-wave measurements of better than 0·95 could be made and repeated. A measurement of the pattern taken with a short-circuit on the end of the guide showed variations between the maxima of not more than ±1%.

A variant of the design was one in which the waveguide for the detector was carried back to the centre pivot, into which was incorporated a rotating joint. This allowed the detector, placed on the remote end of the rotating joint, to remain fixed while the probe was moved. Since the actual rotation angle is only a few degrees, any variations caused by rotation of the joint were unnoticeable. This construction was found useful in the early work of the programme, when oscillator power and detector sensitivity were so small that it was necessary to use a superheterodyne receiver as detector and this could not readily be moved.

(3.2) Frequency

Since much of the work was carried out by harmonic generation from oscillators of around 1·2 cm wavelength, cavity wavemeters were produced for frequency measurements in both the 6 mm and 1·2 cm regions.

Both H_{11}- and H_{01}-type cavity wavemeters were made for these bands, having copper cavities with a plunger controlled by micrometer movement. They were coupled to two waveguides by coupling holes in the sides and could be used as absorption or transmission types. The H_{01} 6 mm type used the coupling method described by Bleaney and Penrose,[8] in which both guides are coupled in at one end of the cavity, the exciting one having

---

* This Committee later revised the standard to 0·188 × 0·094 in, to accord with a standard adopted subsequently by the Radio Manufacturers Association of the United States.

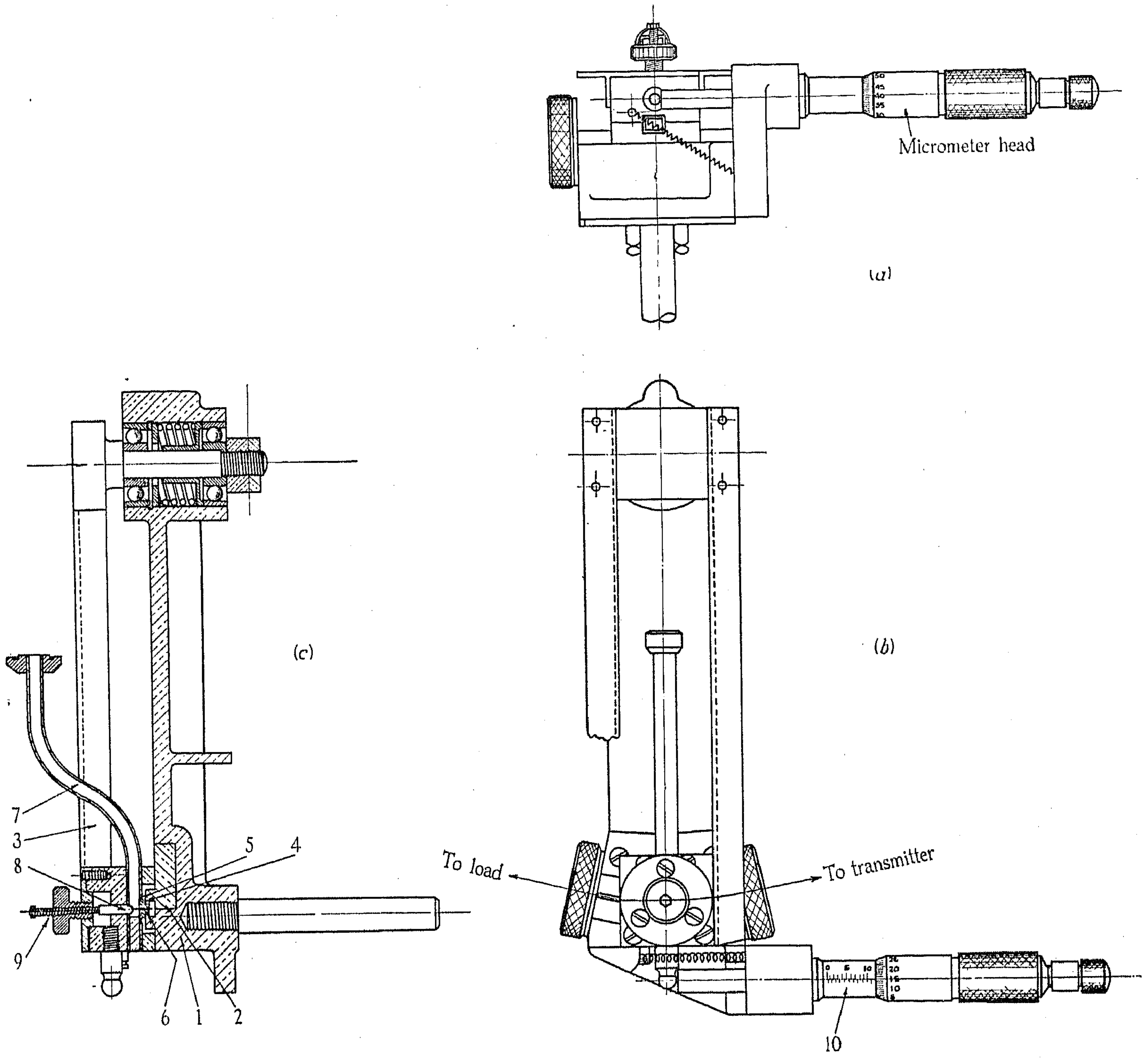


Fig. 5.—Standing-wave detector (WG24).

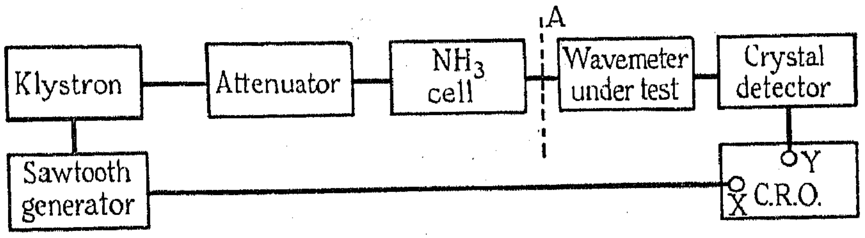


Fig. 6.—Wavemeter calibrating circuit using selective absorption of low-pressure ammonia.

a two-hole coupling and the detector a single hole on a radius at 45° to the line of the other two.

Little trouble from spurious modes was found on the H_{11} types, but the H_{01} types with their larger diameter showed several other modes. The higher Q-factor of the desired mode served to distinguish it without difficulty.

The Q-factors of the H_{11} and H_{01} types were respectively about 3 000 and 18 000 at 1·2cm wavelength, while at 6mm the values were about 1 200 and 7 000 respectively.

At first the only calibrations possible were by calculation from the measured bores of the cavities, but later experimental work[9] on ammonia absorption lines enabled very accurate checks to be made upon the wavemeters for both bands. The ammonia lines lie in the 1·2cm region, and most of them have been determined in frequency to one or two parts in 10^6. The method of calibration is the simple one illustrated in the block diagram of Fig. 6. A sawtooth generator is used to sweep the frequency of a klystron oscillator over a range of about 5Mc/s on a mean frequency around 24 000Mc/s. The power passes into a length of waveguide containing ammonia at low pressure (10^{-2} to 10^{-3}mm Hg) and thence through the wavemeter under test to a crystal detector, whose output is indicated as a function of frequency sweep on a cathode-ray oscilloscope. It is then easy to locate on the oscillograph screen an absorption due to an ammonia line, and set the absorption due to the wavemeter resonance to coincide with it. A setting accuracy of one or two parts in 10^5 is readily attained. With an approximate knowledge of the frequency it is not difficult to identify positively the different lines. Fig. 7 shows the calibration curve of an H_{01} type wavemeter as determined from its dimensions. The 28 ammonia absorption lines which lie between 22 600Mc/s and 25 700Mc/s are inserted as calibration points. The measured diameters have agreed to within 0·0001–0·0002in with those calculated from the calibration points.

For calibration of 6mm wavemeters it was sufficient to insert a second-harmonic generator at the point A in Fig. 6.

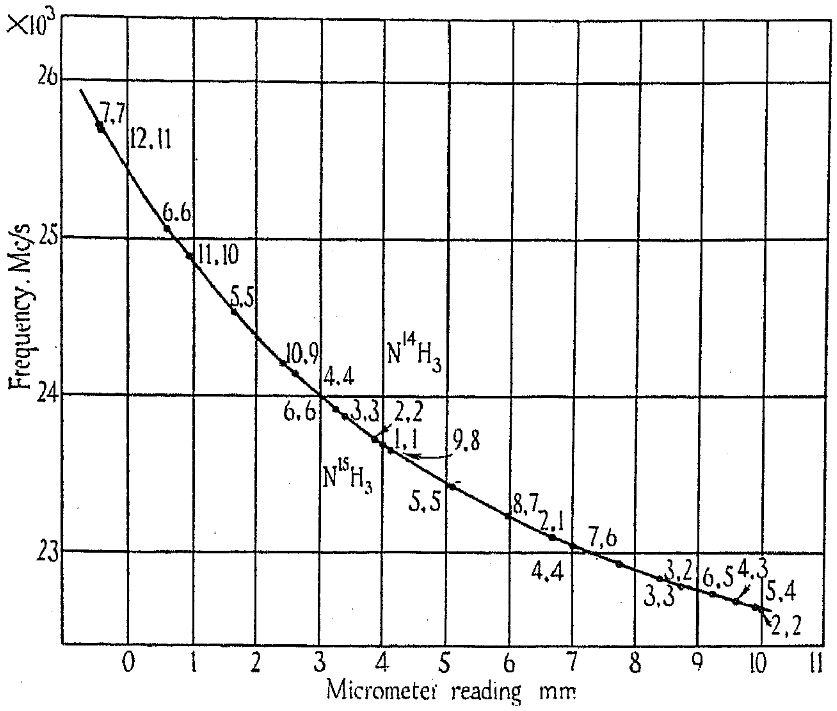


Fig. 7.—Calibration of H_{01} wavemeter showing detected absorptions due to $N^{14}H_3$ and $N^{15}H_3$ isotopes.

(3.3) Power

(3.3.1) High-Power Measurements.

Powers of the order of 0·5 watt and upwards were measured by means of the water-load shown in Fig. 8(*a*), which was scaled down from a design for 3 cm wavelength.[10] The water tube is made of quartz tube of 1 mm diameter and 0·5 mm bore. To give a sharp taper, two lengths of tubing were ground at an appropriate angle and the faces butt-joined. The tube lies in the waveguide, and it is connected to a 12-section thermo-junction unit by short rubber tubes.

When used with a sensitive galvanometer and a water flow of 1·5 cm³/sec, powers of the order of 25 mW can be measured. Response is rapid as there is only a relatively small quantity of water in the tube. An input standing-wave ratio of 0·7 was measured.

(3.3.2) Low-Power Measurements.

The methods of measuring low powers at centimetre wavelengths by means of bolometers, using thermistors or fine-wire elements, are well known. The difficulties at millimetre wavelengths are again those associated with the size of the elements, and the estimation of the errors involved.

A bolometer mounting was made in the form shown in Fig. 8(*b*). This could be used with either a wire filament or a thermistor bead. The element was mounted so that it passed through the centre of the cross-section of the main waveguides, along the direction of the electric field. It then passed through a subsidiary waveguide on either side of the main guide, and was terminated on the outside walls of these guides. Pistons in the subsidiary guides served to vary the effective length of the filament, and a further piston in the main guide provided the main matching control. With the very small powers available in the early part of the work it was found extremely difficult to set these bolometers to a matched condition, and they could be used only as indicators of the orders of power.

When the c.w. magnetron became available, however, it was possible to tune the bolometers satisfactorily. Calibration against the water load previously described at a power level of about 25 mW showed that the indicated power of a thermistor bolometer was about 45% of that indicated by the water load. This figure is somewhat higher than would be expected from extrapolation of the measurements of Collard, Nicoll and Lines,[11] from whose work a figure of little more than 20% would be expected. There seems no particular reason to doubt the accuracy of the higher figure.

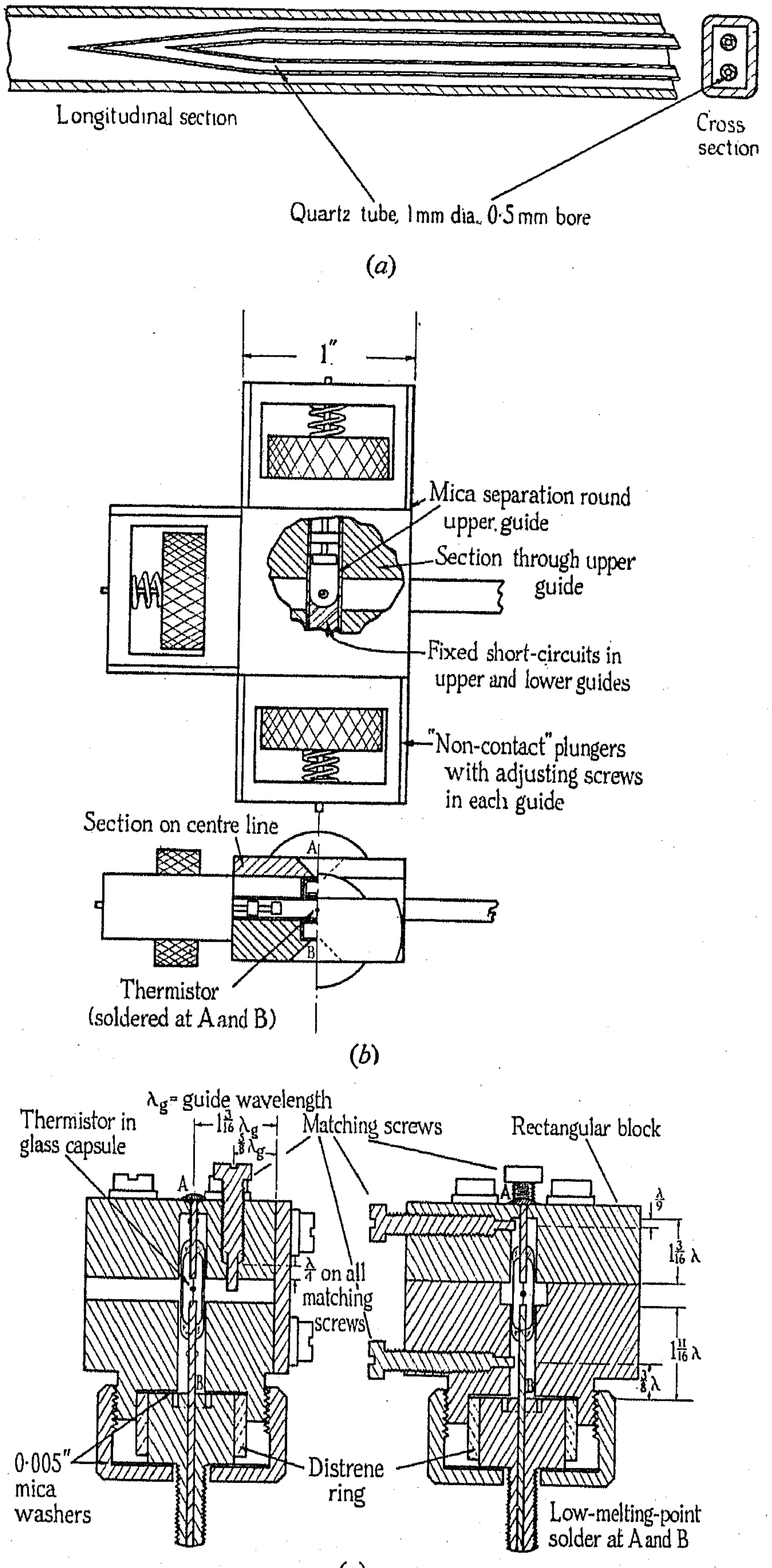

Longitudinal section
Cross section
Quartz tube, 1 mm dia., 0·5 mm bore
(a)
1″
Mica separation round upper guide
Section through upper guide
Fixed short-circuits in upper and lower guides
"Non-contact" plungers with adjusting screws in each guide
Section on centre line
A
B
Thermistor (soldered at A and B)
(b)
λg = guide wavelength
Thermistor in glass capsule
1 1/16 λg
3/8 λg
Matching screws
Rectangular block
A
λ/4 on all matching screws
B
λ/9
1 1/16 λ
1 11/16 λ
3/8 λ
0·005″ mica washers
Distrene ring
Low-melting-point solder at A and B
(c)

Fig. 8.—Power measuring circuits.
(*a*) Water load.
(*b*) Tunable thermistor mount.
(*c*) Pretuned thermistor mount.

Another form of thermistor mounting is shown in Fig. 8(*c*). The waveguide and coaxial arms were short-circuited at predetermined lengths, and matching was carried out by reactance screws in the waveguide and coaxial arms. With the dimensions chosen, the input standing-wave ratio could not be made better than about 0·35, but useful measurements have been made, and indicated powers of around 80 μW have been measured from second-harmonic generators. Comparison with the water load again showed that the indicated power was 45–50% of that measured by the water load.

Some work was also done on metal-film bolometers. A metal film placed across a waveguide should absorb all the incident energy, provided its surface resistivity is equal to the wave impedance of the guide. As the thickness of such a film is considerably less than the penetration depth of currents at these frequencies, it was hoped that the film bolometer would give a more accurate reading than the filament or thermistor type, in which the relevant thicknesses are considerably greater than the penetration depth. Some success was achieved with films of bismuth about 0·01 micron thick deposited on a film of collodion 1 micron thick. The main difficulty was in the rapid deterioration of the unprotected film. Protective coatings of silica and magnesium fluoride were tried without much success. There is little doubt that the techniques now available for the production and protection of the metal films used as attenuators at centimetre wavelengths could be applied successfully to film bolometers. It is felt that the film bolometer is a promising development for measurements at millimetre wavelengths, and would repay further attention.

(3.4) Receiver Sensitivity

The signal generator developed for receiver measurements at millimetre wavelengths used a harmonic generator as described earlier in this paper. The klystron source and harmonic generator were both enclosed in a circular screening tank of the type previously developed for centimetre-wave signal generators.[12] Power leads entered the generator through lossy coaxial lines having a copper-sulphate jelly dielectric. Forced air cooling was required for the klystron, and this was provided by a small air blower which drew in air through a number of holes in the screening tank. These holes were sufficiently small to prevent propagation of the harmonic frequency. Their ratio of length to diameter was about 5 : 1, thus giving an attenuation in excess of 100 dB. Over 100 holes were used to reduce resistance to the air.

The power output of the harmonic source was so small that it was not desirable to include a piston attenuator as an integral part of the signal generator, since its minimum attenuation would be so great that it would be impossible to get any indication on a detector or bolometer connected to its output.

A variable attenuator for external connection to the signal generator is shown in Fig. 9. It is of the piston type, and is so constructed that no relative movement of the input and output waveguide couplings occurs. The attenuating tube has an internal diameter of 2·5 mm, and is fed in the E_{01} mode by means of a disc coupled to the input conductor which projects into the attenuating tube, forming a coaxial line. Alteration of the distance of penetration varies the amount of attenuation. The far end of this conductor serves as a probe to couple it to the output waveguide, which is bent in spiral form to allow movement. The attenuator is controlled by a micrometer drive. Measurements with a standing-wave detector of the input impedance and attenuation of the spiral guide showed no measurable change when the spiral was expanded or compressed.

With a bore diameter of 2·5 mm the attenuation rate at 6 mm wavelength is 14·03 dB/mm for the E_{01} mode. A frequency correction, included in the above figure, is necessary at these wavelengths, since the bore is relatively large.

The law of the attenuator was measured by means of a superheterodyne receiver in terms of the i.f. signal at 45 Mc/s. The curve indicated the presence of a certain amount of the H_{11} mode, which has an attenuation rate of 8·8 dB/mm. Redesign would be necessary to remove the asymmetry to which the presence of this mode must be due, but the attenuator was considered satisfactory for the work required.

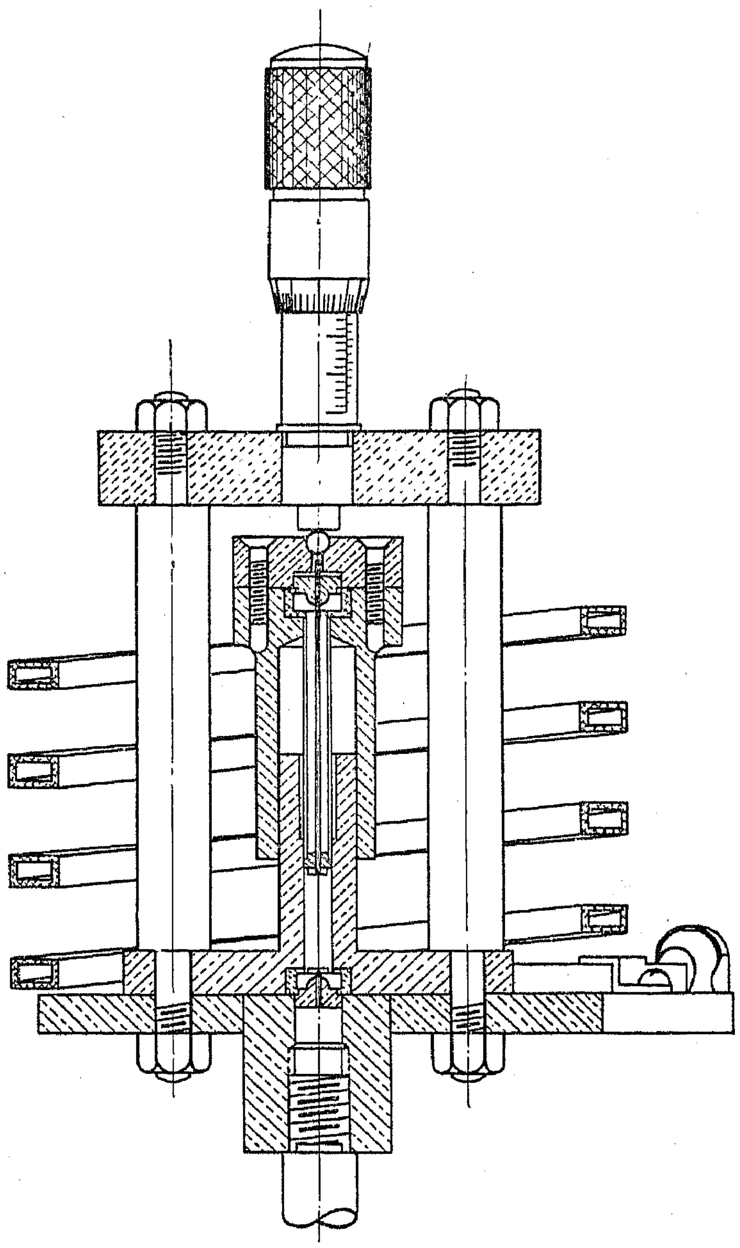


Fig. 9.—Waveguide attenuator (WG24).

(4) TRANSMITTING EQUIPMENT

The propagation measurements to be described fall into three main parts. The first two dealt with two aspects of propagation at millimetre wavelengths—the interference field due to reflection from the earth, and the attenuation due to atmospheric absorption. These were measurements of field strength, and were carried out with a low-power tone-modulated transmitter. The third series of tests was designed to examine the effect of the first two upon a communication system, and used a high-power pulsed-magnetron transmitter with speech modulation.

(4.1) Low-Power Transmitter

The low-power transmitter used a klystron and crystal distorter as described in Section 2.1.1. Square-wave modulation of 1 000 c/s repetition rate was applied to the modulator electrode of the klystron, this being a most useful form of modulation for measurement of field strength by a receiver. A wavemeter was included between the valve and distorter crystal. These components were mounted behind a 16 in-diameter aluminium paraboloidal reflector, which was fed from the front by a waveguide flared at its end. The gain of this aerial was approximately 40 dB at 6 mm wavelength, and its beam width between half-power points was just over 1° in both planes. The transmitter was not directly monitored, because of the difficulty of finding a satisfactory form of monitor for a radiated power of 10–50 μW, which was the normal value for stable output. A meter indicating the crystal current to the distorter was a useful check on any changes taking place in the crystal characteristic.

(4.2) High-Power Transmitter

The high-power transmitter used the pulse-modulated magnetron described in Section 2.1.2, the c.w. version being a much later development. For use with speech modulation some form of pulse modulator was required. The choice lay between modulation of the phase, frequency or width of the pulses, and was decided by the following considerations. The magnetron was designed for operation at a duty cycle of 1/1 000. In any pulsed system the mean repetition rate of the pulses must be well above the highest audio frequency in the signal, and this, coupled with the low-duty cycle, limited the pulse duration to a fraction of a microsecond. Width modulation of such narrow pulses would have required rather complicated circuits, and this method was therefore ruled out. Pulse-frequency modulation was chosen for the first modulator; a later version used pulse-phase modulation.

For speech communication with reasonable intelligibility a frequency range between about 200 and 2 500c/s must be transmitted. With a pulse system the intelligibility improves as the mean repetition rate of the pulses is increased, and in this case a repetition rate of 10kc/s was chosen, this being four times the highest speech frequency to be passed. This required a pulse duration of 0·1 microsec for a duty cycle of 1/1 000. The deviation was arranged to be about ±2kc/s.

The modulator was designed to provide pulses of the order of 15kV 20amp peak, and a hard-valve circuit was chosen to give longer life and freedom from jitter in firing.

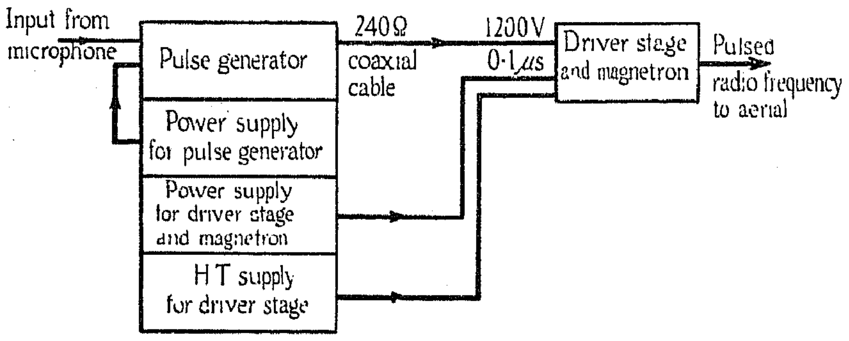
A block diagram of the complete modulator is shown in Fig. 10(*a*). The pulse generator comprised a number of units, as shown in the block diagram of Fig. 10(*b*). The multivibrator generated positive pulses of approximately 100V peak and 10 microsec duration at the mean repetition rate of 10kc/s. This frequency could be varied linearly over the range 8–12kc/s by a modulating voltage from a microphone or from an internal 1kc/s oscillator. A low-pass filter in the microphone circuit removed frequencies above 2 500c/s. The modulated pulse output from the multivibrator was fed into three cascaded amplifier stages having as anode loads short-circuited delay lines with time delays of 1, 0·3 and 0·1 microsec respectively. These transformed the relatively long multivibrator pulse into a 0·1 microsec pulse, and amplified it to a level of about 1 300 volts. A cathode-follower stage reduced the output impedance to 240 ohms in order to match the 240-ohm coaxial cable which transmitted the pulses to the grid of the VX7001 modulator valve. The modulator valve was normally biased off, but on receiving a pulse became conductive and allowed the condenser which coupled it to the magnetron to discharge through the magnetron. The condenser was recharged in the interval between pulses to very nearly the h.t. supply voltage. The pulse as observed on a resistive load had rise-and-fall times each approximately of 0·05 microsec, and a duration at the peak of 0·1 microsec.

The final stage of the modulator and the magnetron were mounted directly on the aerial, which was either the paraboloid described in Section 4.1 or a rectangular horn incorporating a dielectric lens in its aperture. The horn had a length of 10in and an aperture of 12in × 12in. The measured gain was 41 dB, with beam angles between half-power points of 1·2° and 1·5°, parallel and perpendicular to the plane of polarization respectively.

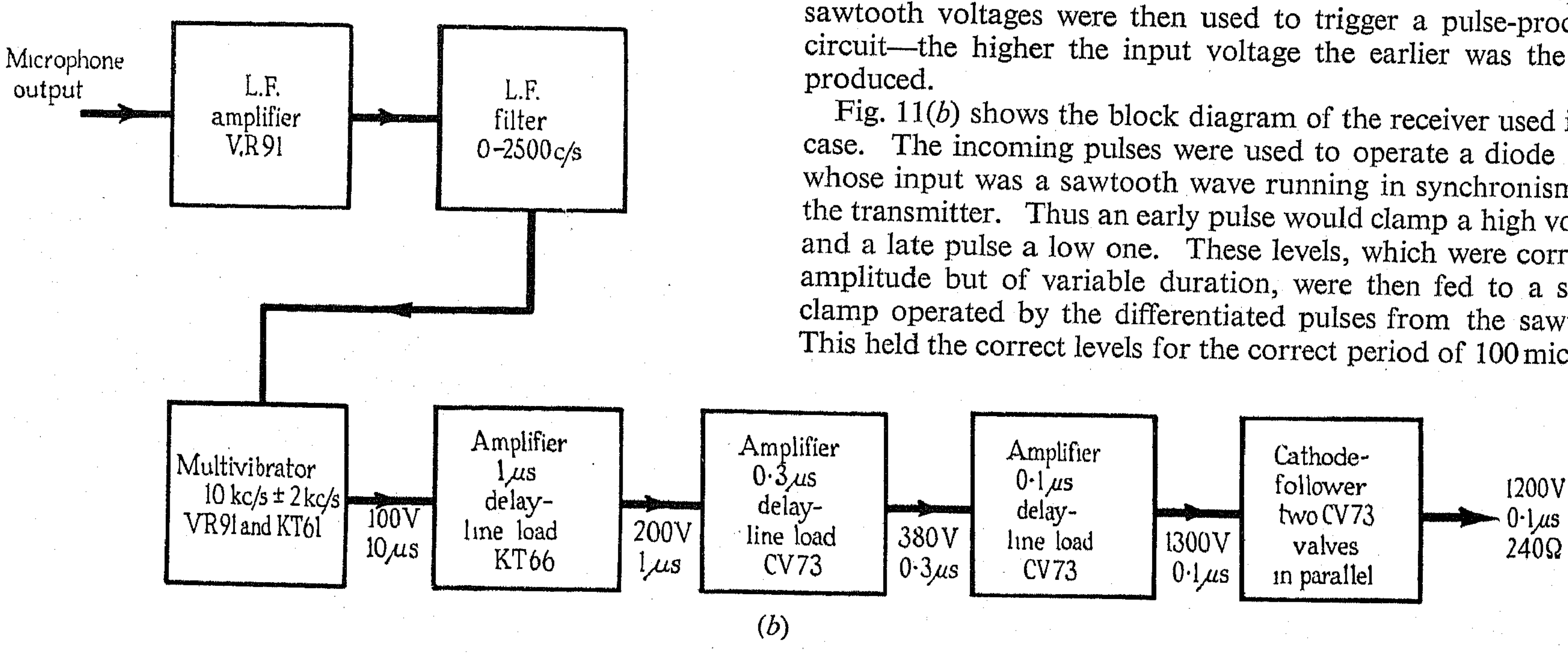
A pulse-phase modulator was also constructed with the aim of increasing the effective modulation depth of the transmitter and thus improving the signal/noise ratio at the receiver. With a mean repetition rate of 10kc/s (100 microsec interval between pulses), the pulse position (phase) could be varied up to ±40 microsec from its mean position to indicate the level of the modulating signal. At the receiver the deviation of these pulses was compared with a locally generated waveform synchronized with the transmitter and used to rebuild the modulating levels.

A block diagram of the transmitter is shown in Fig. 11(*a*). The differentiated 10kc/s sawtooth voltage was used to operate a diode clamp, which cut the a.f. input into a series of steps, each step representing the level of the a.f. waveform at the beginning of a 100-microsec period. The combined step and sawtooth voltages were then used to trigger a pulse-producing circuit—the higher the input voltage the earlier was the pulse produced.

Fig. 11(*b*) shows the block diagram of the receiver used in this case. The incoming pulses were used to operate a diode clamp whose input was a sawtooth wave running in synchronism with the transmitter. Thus an early pulse would clamp a high voltage, and a late pulse a low one. These levels, which were correct in amplitude but of variable duration, were then fed to a second clamp operated by the differentiated pulses from the sawtooth. This held the correct levels for the correct period of 100 microsec.


Input from microphone
Pulse generator
240Ω coaxial cable
1200V 0·1μs
Driver stage and magnetron
Pulsed radio frequency to aerial
Power supply for pulse generator
Power supply for driver stage and magnetron
H T supply for driver stage

(*a*)


Microphone output
L.F. amplifier V.R91
L.F. filter 0-2500c/s
Multivibrator 10 kc/s ± 2kc/s VR91 and KT61
100V 10μs
Amplifier 1μs delay-line load KT66
200V 1μs
Amplifier 0·3μs delay-line load CV73
380V 0·3μs
Amplifier 0·1μs delay-line load CV73
1300V 0·1μs
Cathode-follower two CV73 valves in parallel
1200V 0·1μs 240Ω

(*b*)

Fig. 10.—(*a*) Block diagram of transportable modulator.
(*b*) Block diagram of pulse-generator circuit.

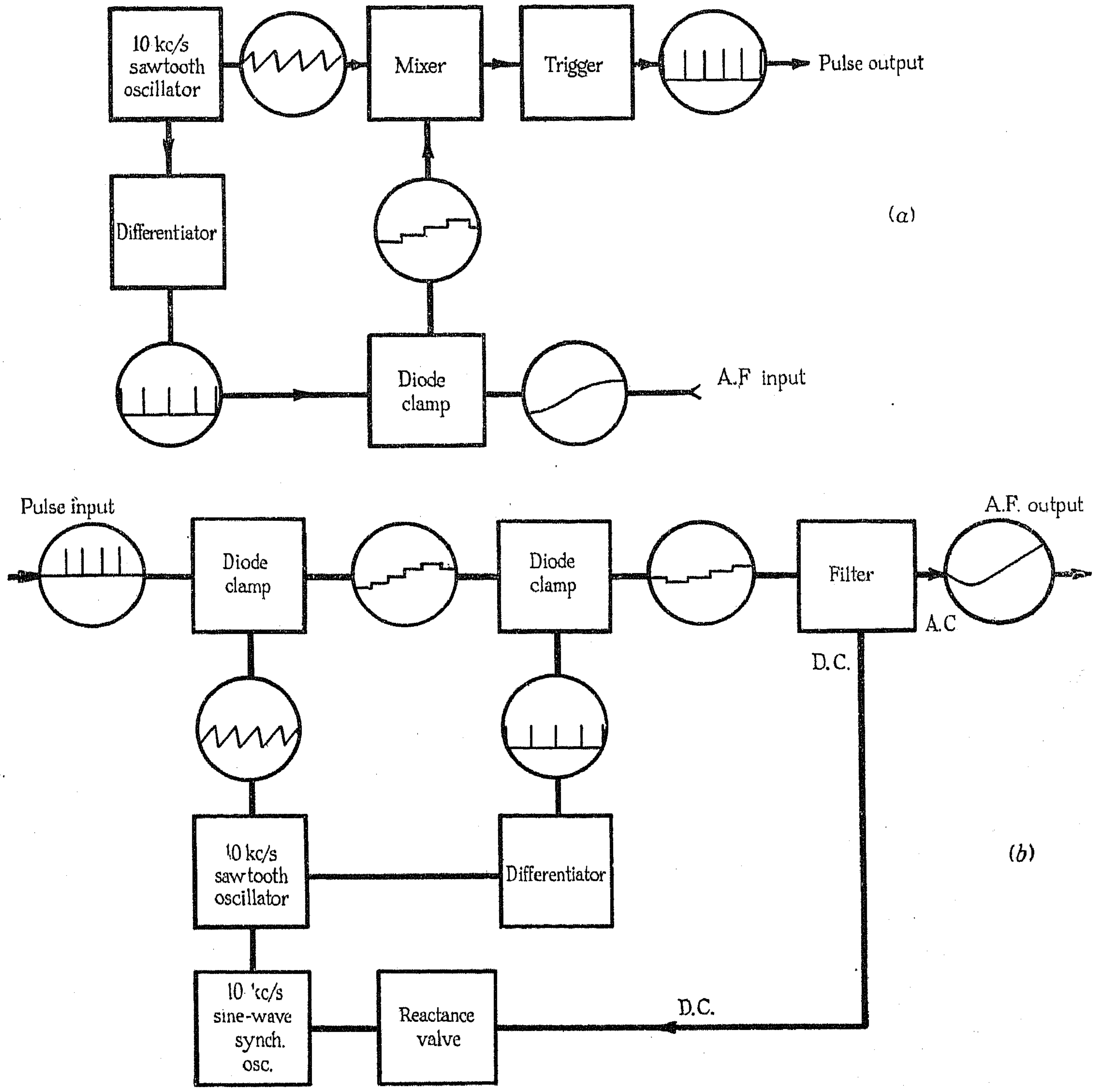


Fig. 11.—(*a*) Pulse-phase modulator.
(*b*) Pulse-phase demodulator.

In this way the original step-wave was built up at the receiver, although retarded in time by 100microsec, and after passing through a low-pass filter the input a.f. waveform was reproduced. This final waveform had a d.c. as well as an a.c. component. If perfect synchronism were maintained between transmitter and receiver sawtooth waves, the d.c. component would remain constant, but should the receiver sawtooth drift, this component would rise or fall (owing to the change in the mean point of intersection of the incoming pulses with the locally generated sawtooth), and use was made of this voltage to correct the phase of the local sawtooth to maintain synchronism.

(5) RECEIVING EQUIPMENT

The receiver used in conjunction with the low-power transmitter was a superheterodyne, having as mixer a capsule crystal as described in Section 2.2. This crystal fitted into an E_{020} resonator. The local-oscillator power, being at a frequency of approximately half the signal frequency, was fed to the cavity in a waveguide channel of suitable size, which was tapered to meet the crystal cavity in an iris coupling. An arrangement which is preferable in theory is that in which the local-oscillator power is fed first to a distorting crystal, and the second-harmonic output taken to an adjacent mixer crystal. In the former method there will be a loss in signal/noise ratio of the order of 10dB over the second method, which has virtually a local oscillator at signal frequency. In practice, however, it was found that, with the small amount of second-harmonic power available, it was difficult to get adequate mixer-crystal drive without undue loss of signal into the distorter crystal. For this reason the single-cavity method was found preferable.

The i.f. output from the crystal was taken through a filter in the base of the cavity to a miniature i.f. amplifier, of mid-band frequency 45Mc/s and bandwidth 15Mc/s, mounted close to the mixer. The complete receiver was mounted on the rear of a 16in paraboloid aerial as used for the low-power transmitter.

Measurements of the noise factor were made, using the signal generator, bolometer and attenuator previously described. Noise factors of approximately 37dB were found at 5·5mm wavelength. Because of the unknown accuracy of the bolometer readings no great accuracy is claimed for these figures, but since in any case there are wide differences between crystals, they can probably be taken as representative of the performance in this frequency band at that stage of development. Advances in technique since the date of the work described here have resulted in a substantial improvement in this figure.

The low-power transmitter was found to have a particular property which could be used to provide a simple but very effective automatic frequency control at the receiver. The transmitter was modulated, as has been stated, by a 1 000-c/s square wave. It was found that during the on periods the frequency changed in a fairly uniform manner by about 5–10 Mc/s. Thus the received signal swept through part of the i.f. frequency 1 000 times per second. Modulation output waveforms from the receiver are shown as A, B and C in Fig. 12(*a*). When the receiver was tuned correctly the output waveform was undistorted as in A, since the excursion of the signal was entirely within the i.f. response, shown in the diagram by a broken line. When the receiver was mistuned this frequency excursion extended to one skirt of the response curve, producing a waveform as shown in B or C, depending on the direction in which the receiver was mistuned.

The effect could be made use of for a.f.c. purposes by deriving a d.c. output whose polarity changed in going from B to C. To do this the circuit of Fig. 12(*b*) was devised. The output waveforms when differentiated by the condenser-resistance combination became A_1, B_1 and C_1. One diode rectified the positive peaks and the other the negative ones, so that the smoothed output was zero, negative or positive and had the usual discriminator characteristic. The output of the discriminator was used to vary the voltage on the grid of a valve whose output controlled the current in the resistance chain from which the reflector voltage of the local oscillator was derived.

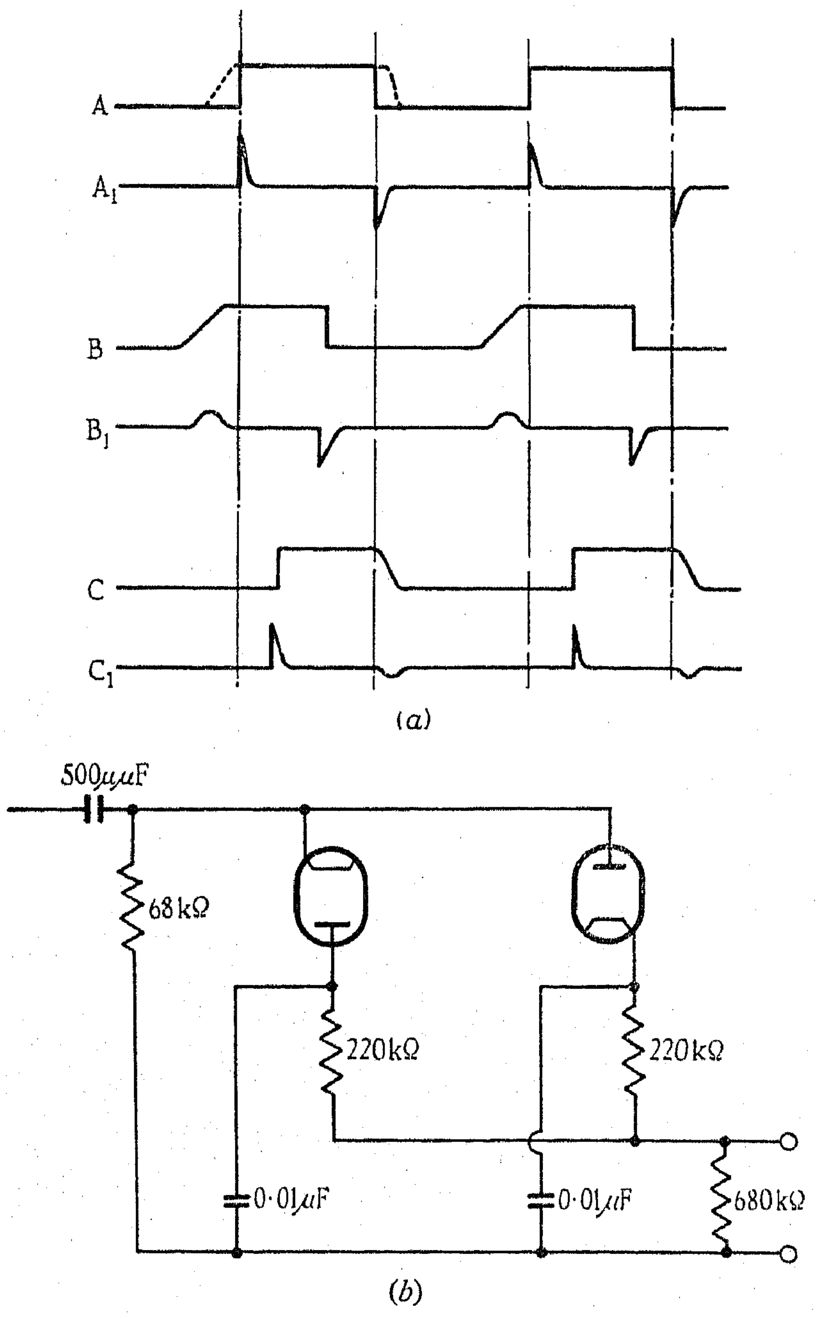


Fig. 12.—A.F.C. system.
(*a*) Waveforms.
(*b*) Circuit.

An advantage of this method was that the output rapidly rose to a maximum for quite small deviations in frequency, producing a very firm control. Another advantage was that it was easy to make and needed no modification to the i.f. amplifier. It needed no readjustment, as the conventional discriminator does, when changing from one channel to the other (i.e. whether the local oscillator was above or below the signal in frequency). It is, however, necessary that the output of the transmitter shall be reasonably constant during the frequency sweep in order to produce a square waveform.

As the receiver was used for measurements of relative signal strength, some form of calibration was necessary. It can be assumed that the i.f. output of the mixer is proportional to the received signal, but small variations in the gain of the i.f. amplifier were troublesome. The difficulty was overcome by feeding a signal, at the mid-band intermediate frequency, to the output side of the mixer through a decoupling resistor. This signal was modulated with a square wave of the same characteristics as the transmitter modulation. It was observed at the receiver output alternately with the signal received through the mixer, the change-over being done by two relays, one of which was a type with high-frequency contacts which formed a switch in the coaxial cable carrying the calibrating signal. The other relay was mounted at the side of the aerial mirror, and through a push-rod operated spring finger which opened or closed over the end of the waveguide feed to the mirror. The output of the signal generator supplying the calibrating signal could readily be adjusted to match the received signal in amplitude. Thus readings were dependent on the accuracy of the attenuator of the i.f. signal generator, and this was calibrated by the National Physical Laboratory at the operating frequency to an overall accuracy of ±0·1 dB.

For reception of pulse-frequency-modulated transmissions no significant modifications to the receiver described above were necessary. The additional circuits used for reception of the pulse-phase-modulated transmission have already been dealt with, in Section 4.

In the third series of field tests the receiving aerial was a biconical horn with an overall diameter of 15 in and a vertical aperture of 6 in, the phase across the aperture being corrected by a polystyrene lens. It was fed by an E_{01} circular waveguide terminating at the apex of one of the cones. The intention was to provide an all-round aerial convenient for use aboard ship. Measurements showed that the azimuth pattern was smooth, with variations up to ±2 dB from the uniform diagram. The vertical beam width was 3° between half-power points.

(6) FREQUENCY STABILIZATION

A further investigation undertaken as part of this programme was in the stabilization of oscillator frequency at millimetre wavelengths. This was desirable for projected long-term measurements on absorption due to rainfall. Stabilization of frequency by means of molecular resonance in gases was an attractive possibility, and the strong resonances in ammonia gas provided very useful standard frequencies. These resonances have already been referred to in Section 3.2. The method developed has been described in detail elsewhere,[9] but an outline is given here with reference to Fig. 13. An oscillator A is swept over a small frequency range on a mean frequency around 25 000 Mc/s. The power passes through a

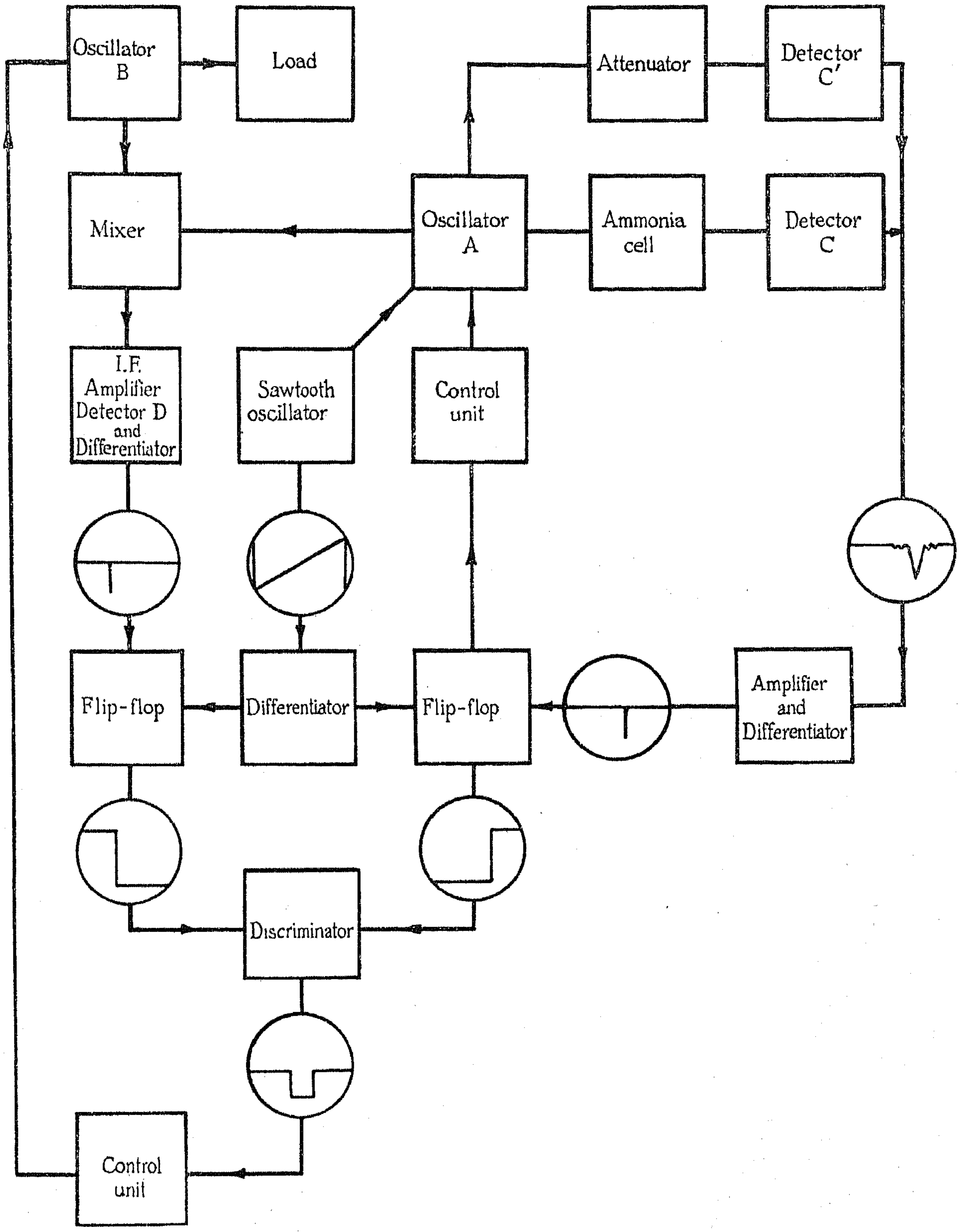


Fig. 13.—Block diagram of oscillator-frequency stabilizer utilizing selective absorption of low-pressure ammonia.

waveguide containing ammonia and thence to a detector C, producing a sharp pulse at the instant when the frequency passes through an absorption frequency. Power from a fixed oscillator B, which is the one to be stabilized, is mixed with power from an oscillator A, and the beat pulse output is fed via a filter to a detector D. A discriminator circuit produces a voltage proportional to the time difference between these two pulses, and this is fed back to control the frequency of oscillator B. A stability better than ±1 part in 10^6 was obtained.

The strong lines in the ammonia spectrum cover a region from about 10·5 to 12·5mm wavelength. Thus they are well suited to stabilizing an oscillator producing harmonic power in the oxygen absorption region, or even an oscillator producing its fundamental power in this region. The strong lines whose second harmonics would lie within the oxygen absorption band are given in Table 1.

The absorption band of oxygen, which is broad at atmospheric pressure, splits up at reduced pressure into sharp discrete absorption lines. Unfortunately, even the strongest of these lines is rather too weak for use in stabilization. The lines listed in Table 2 are some others which have been selected from published

Table 1

Lines in Ammonia Spectrum

NH_3 line J, K*	Line frequency	Second-harmonic wavelength
	Mc/s	mm
6, 6	25 056	5·988
7, 7	25 715	5·827
8, 8	26 519	5·650
9, 9	27 478	5·453
10, 10	28 605	5·238

* J and K are quantum numbers which represent respectively the total number of units of angular momentum, and the number of units of angular momentum about the axis of symmetry (J ⩾ K).

results[13] as of suitable intensity to be used for stabilization within the oxygen band. The superior indices refer to the atomic weights of the particular isotopes, all of which occur naturally. Provided that a sufficient proportion of the required isotope is present, the presence of other isotopes, or indeed or other gases, does not affect the usefulness of the resonance for stabilizing.

Table 2

LINES SUITABLE FOR STABILIZATION

Substance	Wavelength
	mm
Methyl fluoride ($C^{12}H_3F$)	5·87
Carbonyl selenide ($C^{12}O^{16}Se^{80}$) ..	5·33
Carbonyl selenide ($C^{12}O^{16}Se^{78}$) ..	5·30
Carbonyl sulphide ($C^{12}O^{16}S^{32}$) ..	4·93

(7) PROPAGATION

This Section contains a summary of the propagation measurements carried out with the apparatus described above, and some conclusions drawn therefrom.

(7.1) Measurements over Sea

The first set of measurements was made in 1946 over sea with the low-power transmitter mounted on shore and the receiver on board ship. Both aerials were directed by means of telescopes aligned with them. The receiving aerial was stabilized against ship movement by mounting it on a gyroscopically-stabilized platform. Vertical polarization was used. The tests were carried out at two wavelengths, 6·35mm and 5·81mm, the former outside the oxygen band and the latter inside it. Some of the results have been discussed in a previous publication,[14] from which the curves of Fig. 14 are taken. This shows signals at the two wavelengths, the mean of the curve for the longer wavelength following the inverse-distance law. Oxygen attenuation is shown in the other curve by its progressive divergence from the inverse-distance curve. The interference pattern indicates the presence of a strong ray reflected from the sea surface. The broken lines added to the 6·35mm curve are the envelope of the maxima and minima calculated for a reflection coefficient of 0·8, which gives the best fit to the actual curve. This was the highest value of reflection coefficient found during the trials; other values found ranging down to 0·25.

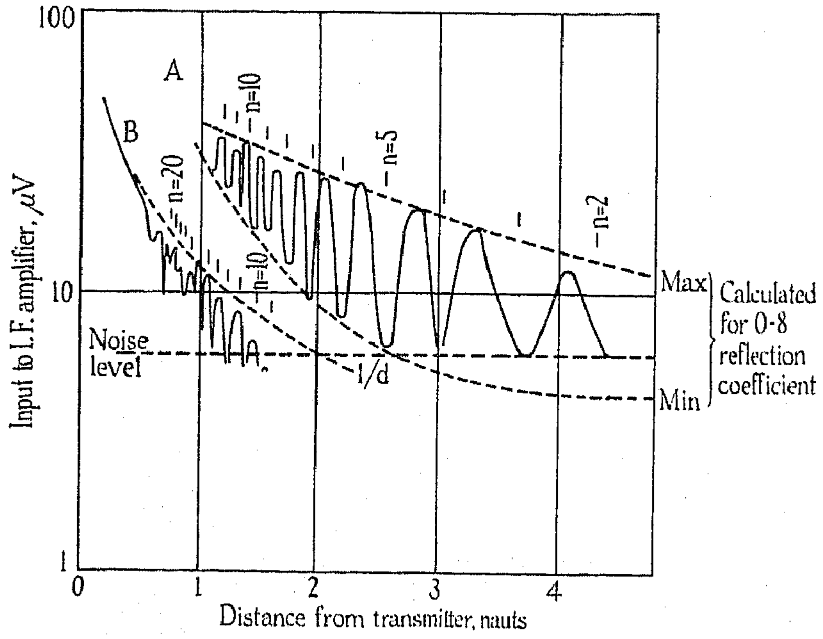


Fig. 14.—Signal-strength/distance curves for millimetre-wave transmission over sea.

A. 6·35mm wavelength.
B. 5·31mm wavelength.
Transmitter height: 15ft.
Receiver height: 63ft.

For the heights used in obtaining the curve of Fig. 14 (15ft for transmitter, 63ft for receiver), the angle of incidence at the sea surface varies from 0·68° to 0·27° between the ranges 2km and 5km. Values of reflection coefficient for a plane sea-water surface, calculated from the known dielectric properties[15] at 10° C, vary between about 0·81 and 0·93 for this range of incidence angle. Thus a value very near to the theoretical was obtained. Even at this wavelength the first Fresnel zone around the reflection point is quite large; for the figures quoted above the major and minor semi-axes of the Fresnel ellipse on the sea surface are 118m and 1·4m for a range of 2km, 462m, and 2·3m for a range of 5km. Thus one would expect to find in practice that for values of reflection coefficient approaching the theoretical the sea surface would have to be smooth over the area of this ellipse. If we use the Rayleigh criterion to distinguish between a smooth and a rough surface, we find that the surface can be considered smooth if surface irregularities have a height h which is less than $\lambda/8 \sin \psi$, ψ being the angle of incidence. This criterion requires h to be less than 7cm at 2km range, and less than 18cm at 5km range. The Rayleigh expression is a rather arbitrary one, and these values may be too large by a factor of 2 or more.

Exact measurements of the sea condition under which the measurements were made were not obtained, but the surface was unbroken, with ripples of the order of 6in and a low swell. Thus it would appear that the conditions, while not definitely those for which completely specular reflection would be expected, did at least approach the required degree of smoothness.

(7.2) Measurements Over Land

One aim of the first series of tests had been to obtain measurements of atmospheric absorption, but it was found that sea reflections and the difficulties of accurate determination of position made adequate measurements impossible. However, as a result of the experience gained, a second series of measurements, made over land in 1947 on short ranges (up to 2km) to avoid reflection, yielded satisfactory absorption figures over a range of wavelengths. For these tests the low-power transmitter was mounted on the control tower of an aerodrome. The receiver, mounted on the roof of a vehicle, was moved to selected sites on the runways. The measurements have been described elsewhere,[16] and Fig. 15 summarizes the results.

In Fig. 15 the measured points are shown in relation to the theoretical curves of Van Vleck for two assumed values of the line-breadth constant $\Delta\nu$, which is an unknown parameter. The values are in obvious agreement with the curve for $\Delta\nu = 0{\cdot}02\text{cm}^{-1}$. The point at the top of the diagram is well off the curve, and it may indicate a line-breadth constant slightly less than $0{\cdot}02\text{cm}^{-1}$, which would increase the height of the line-structure peaks.

Beringer's work[2] indicates a value of the constant somewhere between 0·02 and 0·05, but the spread of his points is considerable. The later work of Strandberg, Meng, and Ingersoll[17] on the microwave spectrum of oxygen at low pressures, in suggesting a value between 0·015 and 0·02, confirms the present authors' conclusions, but from measurements on mixtures of gases Strandberg *et al.* predict attenuations as high as 30–40dB/km in

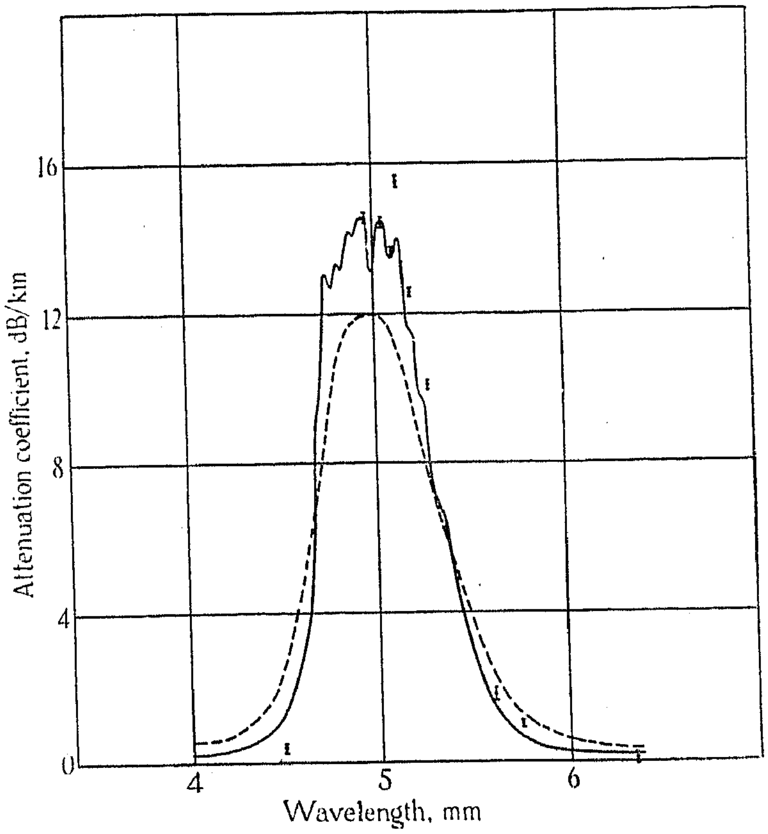


Fig. 15.—Theoretical and measured attenuation due to absorption by atmospheric oxygen.

Theoretical curves.

—— $\Delta\nu/c = 0\cdot02\,\mathrm{cm}^{-1}$.
---- $\Delta\nu/c = 0\cdot05\,\mathrm{cm}^{-1}$.

air. Burkhalter, Anderson, Smith and Gordy,[18] on the other hand, find values of the constant which depend on the particular absorption line and vary from 0·02 to 0·05.

The method used here, although it may not give the information on molecular structure which can be obtained by the techniques of microwave spectroscopy, does nevertheless give a direct measurement of the attenuation actually experienced by waves travelling in the atmosphere.

(7.3) Effect of Propagation on a Communication Circuit

When the high-power transmitter with the pulse magnetron became available, a further short series of tests were carried out in 1947 over sea to investigate communication possibilities. The transmitter was on shore, and had the narrow-beam horn aerial previously described. The receiver was used with both the directional paraboloid and the 360° biconical horn, again fixed on a stabilized mounting. Runs were made to and from the transmitter at a number of wavelengths, and the effects of atmospheric absorption and the interference pattern on the useful ranges were noted at five different wavelengths, between 6·25 mm and 5·39 mm. The attenuation figures agreed very well with those referred to in Section 7.2, and satisfactory speech was transmitted with the directional aerial at ranges up to 5·5 km at 5·39 mm wavelength (where the atmospheric attenuation is 4·8 dB/km) and 13 km at 5·7 mm wavelength (where the atmospheric attenuation is 1·0 dB/km). With the omni-directional aerial at 5·7 mm wavelength a communication range of 4·5 km was reached.

Two observations of interest were made during these tests. First, it was noticed that any object which appeared as a visual obstruction between transmitter and receiver would cause a complete break in signal even at short ranges. Secondly, during one of the later runs rain fell and a substantial loss of signal was noted. The period of rainfall was too short, and its distribution too uncertain, for an accurate assessment of attenuation. However, the incident was of use in giving quantitative support to the calculations of Ryde (in unpublished reports), his figure for moderate rain (4 mm/h) at 5 mm wavelengths being 2 dB/km. It is much regretted that circumstances have not permitted the authors to extend these measurements further.

(8) ACKNOWLEDGMENTS

The authors are indebted to the Board of Admiralty for permission to publish this paper, which describes part of a research and development programme carried out on their behalf.

The work described in this paper was performed between 1945 and 1950 by a number of workers, all of whom made individual contributions to the programme. The task of the authors has been to summarize and correlate these contributions in order to present a coherent account of progress. Those concerned were Messrs. F. Brook, P. C. Butson, J. C. Dix, E. M. Hickin, E. Kettlewell, H. R. L. Lamont, D. A. E. Roberts, R. G. Robertshaw, J. R. Tew, W. E. Willshaw.

The co-operation of the following organizations is gratefully acknowledged:

Admiralty Signal and Radar Establishment, for active co-operation, including the provision of facilities for the propagation trials over sea described in Sections 7.1 and 7.3 and the design of horn and biconical aerials (Sections 4.2 and 5).

Electric and Musical Industries, Ltd., for developing the special velocity-modulation oscillators types VX5015 and VX5016 (Section 2.1.1).

Radar Research Establishment, for the supply of velocity-modulation oscillators (Section 2.1.4).

Royal Aircraft Establishment, for the provision of facilities for the propagation measurements over land (Section 7.2).

(9) REFERENCES

(1) Van Vleck, J. H.: "The Absorption of Microwaves by Oxygen," *Physical Review*, 1947, 71, p. 413.
(2) Beringer, R.: "The Absorption of One-Half-Centimeter Electromagnetic Waves in Oxygen," *ibid.*, 1946, 70, p. 53.
(3) Van Vleck, J. H.: "The Absorption of Microwaves by Uncondensed Water Vapor," *ibid.*, 1947, 71, p. 425.
(4) Megaw, E. C. S.: "Experimental Studies of the Propagation of Very Short Radio Waves," *Journal I.E.E.*, 1946, 93, Part IIIA, p. 79.
(5) Willshaw, W. E., and Robertshaw, R. G.: "The Behaviour of Multiple Circuit Magnetrons in the Neighbourhood of the Critical Anode Voltage," *Proceedings of the Physical Society*, B, 1950, 63, p. 41.
(6) The General Electric Co., Ltd., and Lamont, H. R. L.: British Patent Specification No. 635917.
(7) Hirst, D., and Hogg, R. W.: "Design of Precision Standing-Wave Indicators for Measurements in Waveguides," *Journal I.E.E.*, 1947, 94, Part IIIA, p. 589.
(8) Bleaney, B., and Penrose, R. P.: "Cavity Resonators for Measurements with Centimetre Electromagnetic Waves," *Proceedings of the Physical Society*, 1947, 59, p. 185.
(9) Lamont, H. R. L., and Hickin, E. M.: "The Application of Molecular Resonance to Microwave Frequency Stabilization," *British Journal of Applied Physics*, 1952, 3, p. 182.
(10) Kettlewell, E.: "A Wide-Band Calorimeter for R.F. Power Measurement at 3 cm," *Journal I.E.E.*, 1946, 93, Part IIIA, p. 1407.
(11) Collard, J., Nicoll, G. R., and Lines, A. W.: "Discrepancies in the Measurement of Microwave Power at Wavelengths below 3 cm," *Proceedings of the Physical Society*, B, 1950, 63, p. 215.
(12) Clayton, R. J., Houldin, J. E., Lamont, H. R. L., and Willshaw, W. E.: "Radio Measurements in the Decimetre and Centimetre Wavebands," *Journal I.E.E.*, 1946, 93, Part III, p. 97.
(13) Kisliuk, P., and Townes, C. H.: "Molecular Microwave Spectra Tables," *Journal of Research, National Bureau of Standards*, 1950, 44, p. 611.
(14) Lamont, H. R. L., and Watson, A. G. D.: "Millimetre Wave Propagation," *Nature*, 1946, 158, p. 943.
(15) Saxton, J. A., and Lane, J. A.: "Electrical Properties of Sea Water," *Wireless Engineer*, 1952, 29, p. 269.
(16) Lamont, H. R. L.: "Atmospheric Absorption of Millimetre Waves," *Proceedings of the Physical Society*, 1948, 61, p. 562.
(17) Strandberg, M. W. P., Meng, C. Y., and Ingersoll, J. C.: "The Microwave Absorption Spectrum of Oxygen," *Physical Review*, 1949, 75, p. 1524.
(18) Burkhalter, J. H., Anderson, R. S., Smith, W. V., and Gordy, W.: "The Fine Structure of the Microwave Absorption Spectrum of Oxygen," *ibid.*, 1950, 79, p. 651.

DISCUSSION ON

"A MOVING-COIL RELAY APPLIED TO MODERN HIGH-SPEED PROTECTIVE SYSTEMS"*

NORTH-WESTERN MEASUREMENTS GROUP, AT MANCHESTER, 27TH OCTOBER, 1953

Mr. H. G. Bonson: The system of balanced-earth-fault transformer protection described in Section 4.4 is applicable only to solidly earthed transformers where heavy neutral current can be guaranteed under all fault conditions. Although this is no doubt the most difficult condition for an unbiased relay, a biased scheme of wider application is surely worth consideration.

The system of Fig. 4 with one main current transformer replaced by the three line current transformers, but with an additional small load-bias feature also derived from the three line current transformers, would not suffer from the weakness of failing to trip when the main in-feed for a transformer-winding fault was from the line side.

For the 80–90% line-length setting of impedance protection mentioned at the beginning of Section 6.2 to be safe, it is necessary that the accuracy of the zone-setting shunts, shown in Fig. 13, should be as great as that of the relay. This is difficult to achieve when, say, a 0–25-ohm resistor is being used at less than 10 ohms, and in my experience inaccuracies of ±10% are to be expected. The solution appears to be either (*a*) the provision of a fixed resistor of, say, 5 ohms resistance in the case mentioned, in series with the variable resistor, in order to reduce the overall error produced by a given percentage error in the variable resistor, or (*b*) connection of the shunts somewhat in the form of the "universal shunt" arrangement given in textbooks.

From the time/current curve of Fig. 2 it appears that an impedance relay operating on a fault just inside its setting distance might take 0·5sec or more to trip. Such a long clearance time carries the risk of incorrect discrimination between adjacent impedance-protected feeders, when the normal time interval between zones of 0·4–0·5sec is employed. Do the authors agree that it may be advisable to increase the zone-3 time-relay setting to, say, 1·2sec?

Mr. H. Easton: The moving-coil relay used by the authors is quite a large one for operating a single contact, but I am certain that these generous proportions are fully justified. Some gear is constantly being reduced in size, since it is remarkable, for instance, what one is expected to house in miniature measuring-instrument cases. When dealing with protective relays, however, this modern tendency should be resisted. A relay with such important work to do should not be handicapped by discs, magnets and similar parts which are too small for the performance expected from them.

It is unfortunate that the rectifier characteristics are such that auxiliary current transformers are needed, since this is a further complication which must be reflected in the cost. It is not to be expected that the authors' versatile relay system will, in fact, replace existing simple relays where these are now satisfactory, but instead it creates new possibilities very useful in these days of expansion in the electrical power industry.

Mr. R. B. Haworth: Referring to Section 4.2, which describes an overall biased differential scheme for power transformers, the authors employ an instantaneous relay of the standard attracted-armature pattern when quick clearance times are essential, as with high internal fault currents. The time of operation of the moving-coil relay at ten times the setting current (approximately the setting required for the instantaneous relay) is 35millisec, as seen from Fig. 2. It does not appear to be an economic proposition to employ an instantaneous relay in order to obtain such a small reduction in time as 15millisec, i.e. from 35 to 20millisec. Could the authors comment on this point and also indicate whether the instantaneous relay is installed to overcome difficulties encountered owing to the effect on the moving-coil restraint element of third-harmonic currents produced in the main current transformers under high fault conditions?

Mr. K. F. Whittle: With balanced voltage feeder protection using rented telephone pilot circuits, it is stated in Section 5.2 that continuous pilot-circuit supervision is provided, which reduces the probability of maloperation. This is not entirely true, since it has been found that the circulatng direct current in the pilot circuits is sufficient to cause operatioin of the relay if the starting-relay contacts, as shown in Fig. 12, fail to provide an efficient short-circuit across the operating coil.

In the relay used for impedance protection it has been found that the zone-1 setting resistance is not very accurate at the lower limits. These resistors should be of better quality, so that the settings may be changed with the protection in service, without the necessity of having to carry out further secondary injection tests as a check on the setting.

Messrs. C. Ryder, J. Rushton and F. M. Pearce (*in reply*): We reply to the points raised by individual speakers under separate headings.

Mr. Bonson.—The suggestion relating to Fig. 4 would seem to suffer from lack of bias during through phase faults.

The inherent accuracy of the impedance relay is unaffected by the scale marking on the zone-setting shunts. We have experienced no difficulty in this respect.

Discrimination is determined by the preset margins between the zones. This is not affected by the conditions postulated, and no increase in time setting is required.

Mr. Easton.—The proportions adopted for the new element enable it to be accommodated within the same dimensions as existing single-pole protective relays, and at the same time permit the traditionally robust construction demanded for protective relays.

Mr. Haworth.—The inclusion of a high-set instantaneous element in the transformer-protection relay reduces the operating time to about one-quarter, and in our opinion is justified.

Mr. Whittle.—The primary need for pilot-supervision equipment is to combat interference from otherwise sound pilot wires. Faults on these wires are more likely to occur than the electrical failure of some component of the protective scheme. The supervision equipment must not itself produce risks of maloperation; this has not so far posed insuperable problems.

---

* Paper by C. RYDER, J. RUSHTON and F. M. PEARCE (see 1953, 100, Part II, p. 261).

DIGESTS OF INSTITUTION MONOGRAPHS

VIBRATORY POWER CONVERTORS: AN ANALYSIS OF PERFORMANCE AND DESIGN

621.314.5 Monograph No. 109 RADIO SECTION

R. H. EVANS, B.Sc., Graduate

(DIGEST *of a paper published in September,* 1954, *as an* INSTITUTION MONOGRAPH, *and to be republished in Part C of the* PROCEEDINGS.)

The vibratory convertor is a means of power conversion from direct current to alternating current, or to direct current at a different voltage. The power source is usually of low voltage and low impedance, e.g. a battery.

The vibrator itself consists essentially of a reed carrying contacts and maintained in continuous vibration by electromagnetic means. These contacts operate in conjunction with a transformer to produce a flat-topped alternating voltage waveform of peak value equal to the battery voltage V_1; the transformed voltage may be rectified in a conventional manner, or with a self-rectifying type of vibrator, by means of additional contacts on the reed. Various circuits are shown in Fig. 1.

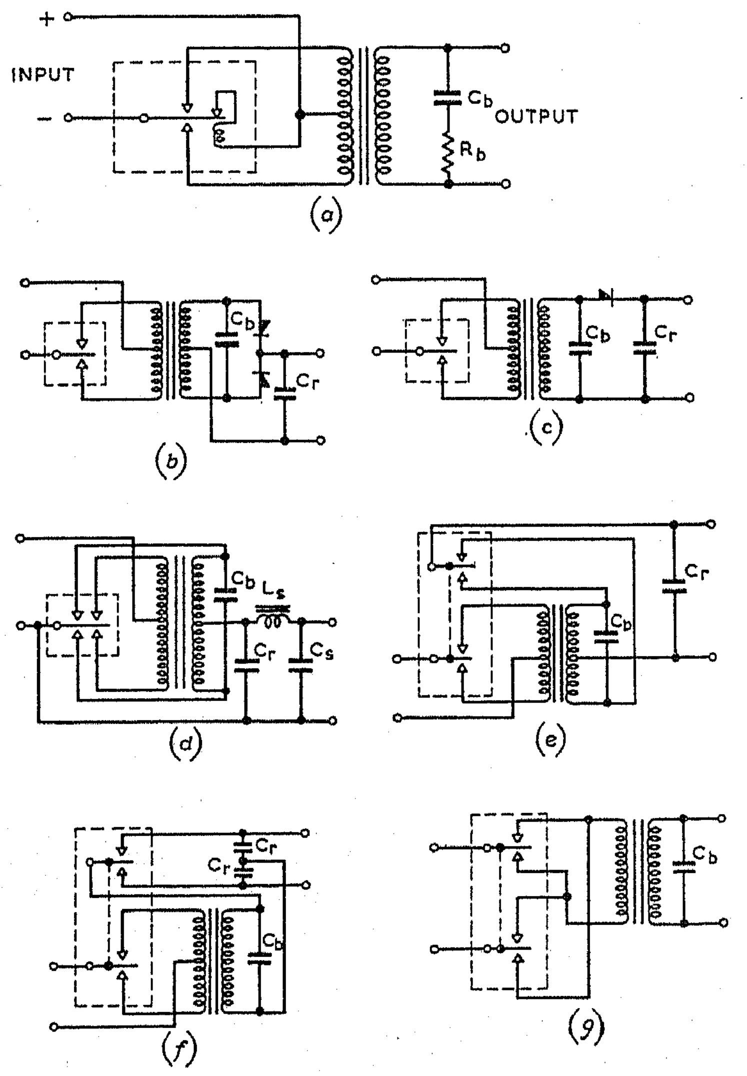


Fig. 1.—Vibratory-convertor circuits.

(*a*) Basic circuit, a.c. output.
(*b*) Interrupter vibrator with separate rectifier, full-wave.
(*c*) As (*b*) but half-wave rectification.
(*d*) Self-rectifying vibrator, full-wave.
(*e*) Split-reed self-rectifying vibrator, full-wave.
(*f*) Split-reed, self-rectifying vibrator, voltage-doubling circuit.
(*g*) Double-pole change-over input, a.c. output.

In the paper a detailed analysis is made of the various circuits and the transformer operating conditions, and formulae are derived for use in design. Where possible, the formulae are non-dimensionalized and displayed graphically.

The spacing of the contacts on either side of the reed is so adjusted that each is closed for about 0·39 of a cycle; this ratio is termed the closure-time ratio, k. Thus in each cycle there are two intervals, each of duration $(1 - 2k)/2$ of a cycle, in which both contacts are open. The buffer capacitor (C_b in Fig. 1) is chosen so that at no load there is an almost linear reversal of voltage during each open interval, i.e. the voltage waveform is trapezium-shaped. Without this capacitor, the sudden interruption of the transformer magnetizing current as the contacts opened would induce high voltages, with consequent destructive arcing. An expression giving the magnitude of buffer capacitance for an ideal waveform is derived in terms of the transformer magnetizing current, which may in turn be obtained from a magnetization curve of the core material and a knowledge of the effective air-gap at the lamination joints. For a given core material, it is shown that the required capacitance is inversely proportional to the square of the e.m.f. induced in the winding across which it is connected and is a function of the ratio k, the peak flux density, the weight of iron, and the ratio of gap length to iron path. It is independent of the vibrator frequency.

In practice it is usual to provide a larger capacitance than the ideal one in order to allow for a fall in closure time as the contacts wear. This "over-buffering" results in a sudden rise in capacitor voltage each time the contacts close, accompanied by a pulse of current of high peak value. The copper loss due to this current constitutes part of the no-load (constant) losses of the transformer, usually amounting to about 20%.

The load characteristic of the convertor depends upon the type of load. If the output is alternating current supplying a resistive load, the buffer capacitor is rapidly discharged each time the vibrator contacts open, so that the voltage and current waveforms are substantially rectangular. The characteristic is then linear, and has the equation

$$V_{2R} = \sqrt{(2k)}nV_1 - RI_{2R}$$

where V_{2R}, I_{2R} are the r.m.s. output voltage and current respectively, n is the transformer turns ratio, and R is the total circuit resistance referred to the output side. With a highly inductive load, such as a fluorescent lamp with its series stabilizing choke, the waveform resembles the trapezium shape typical of no-load conditions, provided that the buffer capacitance is sufficient to supply the load current as well as the transformer magnetizing current during the open intervals of the contacts. The optimum magnitude of this capacitance is calculated, and curves are given for computing the r.m.s. output current from the resistance and inductance of the load and the peak induced secondary e.m.f.

The equivalent circuit referred to the output when a self-rectifying vibrator is employed is shown in Fig. 2(*a*); the transformer leakage inductance has been omitted, since its effect is negligible in the majority of designs for power outputs up to about 50 watts. The peak e.m.f. induced in the secondary

---

Mr. Evans is at the Royal Aircraft Establishment, Farnborough.

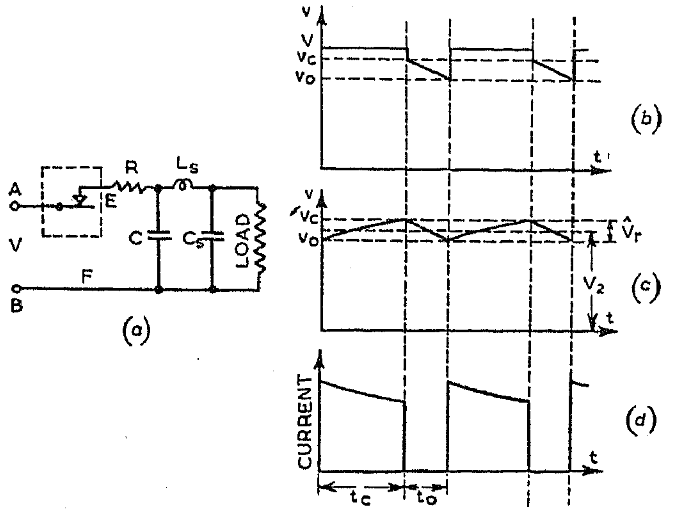


Fig. 2.—Equivalent circuit, d.c. output, neglecting leakage inductance of transformer.

(*a*) Equivalent circuit referred to output.
(*b*) Voltage across EF.
(*c*) Voltage across capacitor C.
(*d*) Current in resistor R.

winding appears across the terminals AB, although it is, of course, effective in supplying the load only during the portions of the cycle in which the rectifying contacts are closed. R is the circuit resistance, C the reservoir capacitance, and C_sL_s constitute the smoothing circuit. If the assumption usual for capacitor-input systems is made that the smoothing inductance L_s is sufficiently large for the current flowing in it to be substantially constant throughout the cycle, it follows that the output characteristic is linear and of the form

$$V_2 = nV_1 - R_eI_2$$

V_2 is the mean rectified voltage across C, and I_2 is the load current. R_e is the effective resistance of the convertor viewed from the output terminals, and may be written sR where s is a multiplying factor; it is shown that s is equal to the square of the form factor of the current flowing in R, and is therefore greater than unity.

Numerical values for s in terms of the circuit parameters are plotted for full-wave, half-wave and voltage-doubling rectification. For large reservoir capacitance the multiplying factor is substantially independent of reservoir capacitance and vibrator frequency, and becomes inversely proportional to the ratio k; this simplified relation may be assumed for most convertors employing full-wave rectification.

Where the transformer leakage inductance is substantial, or extra series inductance is introduced for the purpose of shaping the current pulses, the performance is modified in the following respects:

(*a*) The output voltage for a given reservoir capacitance may be either decreased or increased, depending upon the relative magnitudes of the circuit parameters.

(*b*) If the reservoir capacitance is varied, the output voltage attains a maximum at some finite capacitance instead of at infinite capacitance.

These results are due to a resonance effect between the leakage inductance and the reservoir capacitance; they are similar to those noted in transformer-rectifier systems operating from a sinusoidal supply.[6] The output characteristic is still linear and has the same equation as for the non-inductive case, but the expression for s is more complicated.

The fraction of the total power that is dissipated in the transformer, and hence the permissible power rating of a given size of transformer, depends upon the type of circuit in which it is employed. Numerical comparisons may be made by means of a power-rating coefficient, values of which are computed for vibrator transformers with an a.c. output or employing the various rectifying circuits, with or without a centre-tapped primary winding. It is shown that the power ratings for a given size of vibrator transformer working at a given flux density are about 25% lower than those of transformers operating from a sinusoidal supply of the same frequency.

The peak-to-peak ripple voltage across the rectified output is directly proportional to load current, inversely proportional to reservoir capacitance and vibrator frequency, and substantially independent of the impedance and output voltage of the transformer. For full-wave, half-wave and voltage-doubling rectification respectively and the same circuit parameters, the relative magnitudes of the ripples are 1 : 4 : 2. The fundamental ripple frequency is that of the vibrator for half-wave rectification, and twice this frequency for the other circuits. By means of harmonic analysis the amplitudes of the fundamental and the second harmonic are derived.

Conventional vibratory convertors are not usually rated at more than approximately 50 watts, but special circuits[7] have been described for increasing this rating several fold by substantially reducing the rate of transfer of contact material for a given mean current. This is achieved by so shaping the current pulses that the contacts open at virtually no load. For example, if in a circuit employing a self-rectifying vibrator the reservoir capacitance is progressively decreased from a large value, then provided that the circuit inductance is sufficiently high, the current flowing in either transformer winding during each half-cycle of vibrator operation assumes the waveform illustrated in Fig. 3. Evidently the desired condition of contact unloading occurs at (*c*) and (*f*), although only (*c*) is of practical importance. The required circuit inductance is usually inserted as a choke in series with the transformer secondary winding.

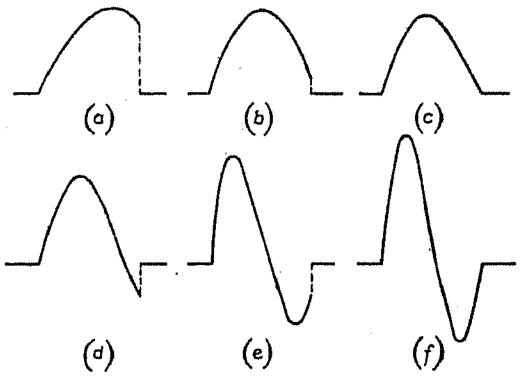


Fig. 3.—Waveforms of current flowing in rectifying contacts during one half-cycle of the vibrator.

(*a*)–(*f*) are for decreasing values of reservoir capacitance; contact unloading occurs at (*c*) and (*f*).

An expression is derived for the necessary relation between the circuit parameters for contact unloading, and the proportioning of these parameters is further discussed with a view to rendering the unloading property independent of variations in the load, the contact timing, or the circuit resistance. It is concluded that the employment of centre-tapped windings should be avoided, that the smoothing inductance and, subject to an upper limit, the reservoir capacitance should each be large, and that the contact surface resistance should be kept low and constant, e.g. by filling the vibrator with a non-oxidizing gas. Expressions are derived for the mean rectified voltage and the ripple voltage in circuits employing contact unloading.

REFERENCES

(1) RADIO COMPONENTS STANDARDIZATION COMMITTEE: "Guide on Vibrators," RCG170, *Radio Components Standardization Committee* (Ministry of Supply, March 1949).

(2) Distin, L. S.: "Modern Vibratory Power Convertors," *Post Office Electrical Engineers' Journal*, 1946, **39**, Part 2, p. 53.
(3) Connelly, F. C.: "Transformers" (Pitman, London, 1950), p. 326.
(4) Mitchell, J. H.: "Recent Developments in Vibrators and Vibrator Power Packs," *Journal of the British Institution of Radio Engineers*, 1952, **12**, p. 431.
(5) Allen, A. L.: "Long-life Contacts for Unidirectional Currents of 1–20 Amperes," *Proceedings I.E.E.*, (Paper No. 1506, July 1953), **100**, Part I, p. 158.
(6) Terman, F. E.: "Radio Engineering," 2nd Edition (McGraw-Hill, New York, 1937), p. 491.
(7) Dixey, K. H., and Wilman, C. V.: "Methods of increasing the Power Rating of Vibratory Convertors," *Proceedings I.E.E.*, (Paper No. 1047 R, March 1951), **98**, Part III, p. 105.

THE A.C. IMPEDANCE OF PLASMA DISCHARGES IN MERCURY VAPOUR

621.3.015.5 : 621.3.011.21 Monograph No. 101

S. E. YUSSUF, Ph.D., Graduate, and Professor J. C. PRESCOTT, D.Eng., Member

(Digest *of a paper published in June,* 1954, *as an* Institution Monograph *and to be republished in Part C of the* Proceedings.)

The path of a direct-current arc discharge in mercury vapour offers an impedance to a superimposed alternating current of small amplitude. This impedance has a resistive and an inductive component. Both components vary with frequency as shown in Figs. 1A and 1B. The dependence of both components on frequency is explained to be due to the time necessary to perform the ionization in the positive column. The purpose of the paper is to derive an expression for the impedance of the column and to study the factors on which it depends. Owing to the large difference in character between the high-pressure and low-pressure arc discharges, each case is considered separately. Some experimental investigations carried out on the low-pressure case are described, and the experimental findings are compared with the theoretical deductions.

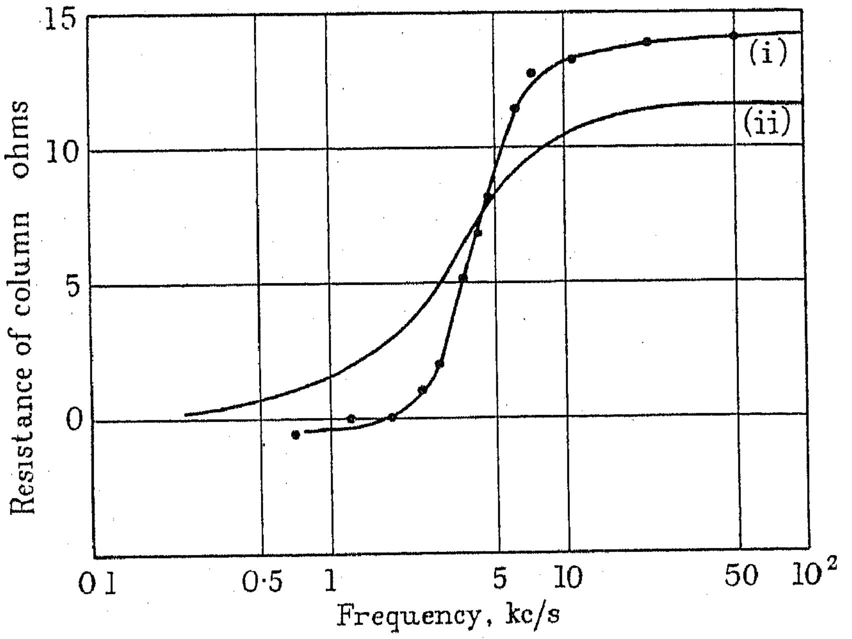


Fig. 1A.—Effective resistance of the 150 cm positive column.
Arc current .. .. .. 4 amp
Bath temperature .. .. .. 40° C
(i) Experimental.
(ii) Calculated.

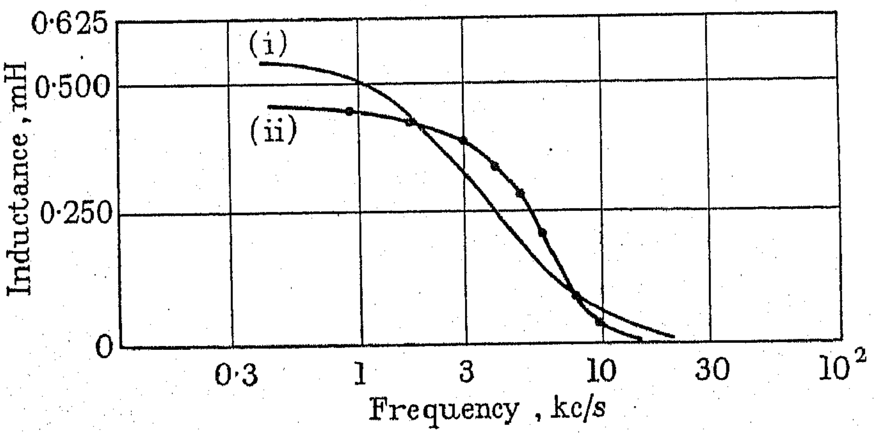


Fig. 1B.—Effective inductance of the 150 cm positive column.
Arc current .. .. .. 4 amp
Bath temperature .. .. .. 40° C
(i) Calculated.
(ii) Experimental.

The High-Pressure Positive Column

When the gas pressure is higher than about half an atmosphere, the electrons, the positive ions and the gas molecules in the column will have the same temperature. The electron concentration was given by Saha and justified by Orstien assuming the particle velocity to be Maxwellian in character and the ionization in the column to be a purely thermal and reversible process. The loss of charged carriers from the column is mainly due to volume recombination. The positive column of the high-pressure arc can be considered to be a core of uniform cross-section and electron concentration. The effective radius and gas temperature of the core can be determined for a certain arc current and electric field. Again the total input power per unit length of the core drawn from the source sustaining the discharge is delivered by the electrons to the gas molecules.

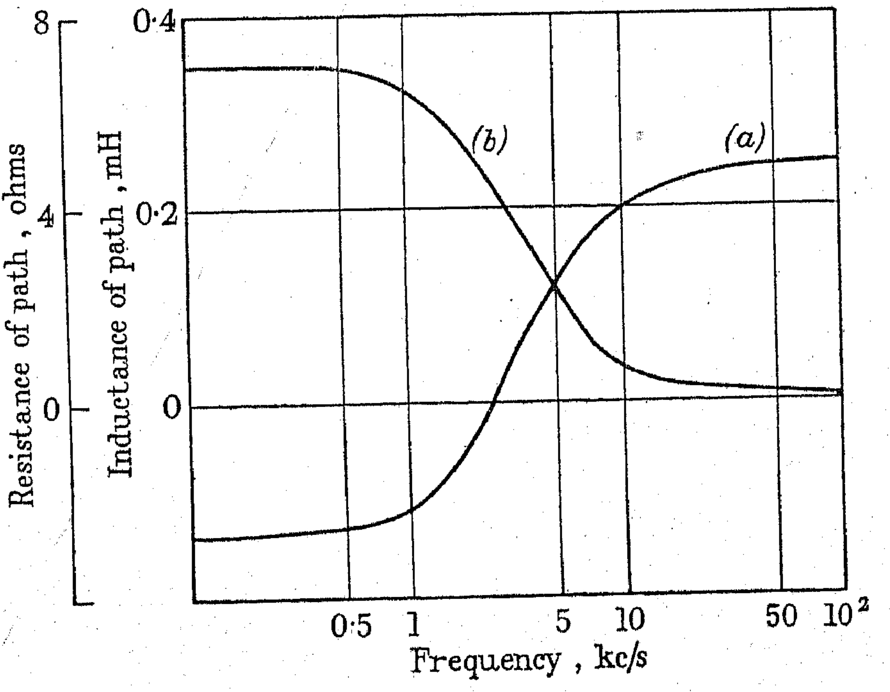


Fig. 2.—Calculated resistance and inductance of a 3 cm 5 amp d.c. positive column at atmospheric pressure.
(a) Resistance.
(b) Inductance.

Dr. Yussuf and Prof. Prescott are at King's College, Newcastle upon Tyne.

This energy escapes by radiation, convection, conduction and by the diffusion of the charged particles carrying their thermal and ionization energy.

Any small changes in the arc current can be attributed to corresponding changes in electron density or changes in the electron drift velocities. It is clear that the former will depend on the rates of diffusion and ionization. However, it can be shown that the effect of diffusion can be neglected if the rate of change of concentration is higher than a certain value beyond which the effect is mainly due to ionization. When the rate of change of electron concentration is small, i.e. at low frequencies, the periodic time is long enough for most of the molecules to take part in the thermal-ionization process producing a large change in the electron concentration. As the frequency is increased, however, the slower molecules will fail to transfer their energy to other molecules. It is therefore to be expected that the change in electron concentration will diminish as the frequency is increased. At very high frequencies no changes in concentration will be encountered, and the arc will behave as a conductor of constant positive resistance. In Fig. 2 are shown the effective resistance and inductance of a positive column in mercury vapour.

The Low-Pressure Case

At low gas pressure the electron temperatures are rather high (2×10^4 °K), and the ionization is performed by electron collisions with the gas molecules. Owing to the high electron temperature, volume recombination is unlikely and the radial diffusion to the walls of the container can account for the loss of charged particles from the column. The electron concentration across the positive column can be defined by a Bessel function with its maximum at the axis of the column, the concentration falling to zero at the walls of the container. The input power per unit length is drawn by the electrons from the source sustaining the discharge. Part of this power is delivered to the gas molecules during the various collisions, and the rest is delivered to the walls as a result of the radial diffusion of the charged particles.

The periodic time of the cycle of variation of the superimposed current will have the same effect on the electron concentration as in the high-pressure case. At low frequencies, most of the electrons can transfer their excess energy to the gas molecules. The number of electrons taking part in the ionization process decreases as the frequency is increased. Again, as the frequency is increased the periodic time will be short for the lower-speed electrons to diffuse to the walls. At high frequency, it will be expected that no changes in electron concentration will occur with the rapidly-changing electric field. The arc path will then behave as a conductor of constant resistance.

Experimental Investigation

The experimental work is designed to measure the effective impedance of the positive column at different frequencies under different conditions of length of column, arc current and vapour pressure. The arc was drawn through either of two 2in-diameter tubes, 150cm and 75cm in length. Each tube was provided with two fixed probes projecting at points along its axis to determine the electron temperature, concentration and axial electric field. In addition a movable probe (see Fig. 3) was provided in the 150cm tube to give the electron concentration at different points along the tube diameter. The mercury cathode pool was placed in a water bath and kept at the required temperature.

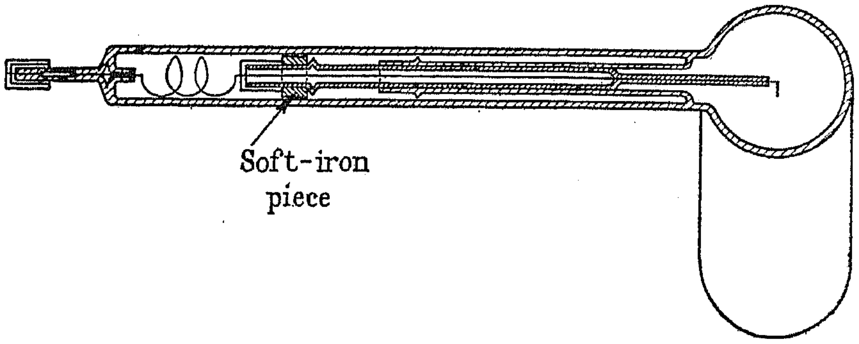


Fig. 3.—Movable probe in the 150cm tube.

A small alternating current of adjustable frequency was injected into the d.c. arc path from a power amplifier. The impedance of the path was then measured on a bridge. Figs. 1A and 1B give the effective resistance and inductance plotted against frequency for an arc current of 4amp in the 150cm tube at 40°C bath temperature.

References

(1) Reich, H. J.: "Theory and Application of Electron Tubes" (McGraw-Hill, 1944).
(2) Von Weizel, Rompe, and Schulz: "Zur theorie der Modulation eines Hochdruckbogens," *Zeitschrift für Technische Physik*, 1940, **12**, p. 387.
(3) Cobine, J. D.: "Gaseous Conductors" (McGraw-Hill, 1941).
(4) Killian, T. J.: "The Uniform Positive Column of an Electric Discharge in Mercury Vapour," *Physical Review*, 1930, **35**, p. 1238.
(5) Langmuir, I., and Mott-Smith, H.: "Studies of Electric Discharges in Gases at Low Pressures," *General Electric Review*, 1924, **27**, pp. 449, 538, 762 and 876.
(6) Maxfield, F. A., and Benedict, R. R.: "Theory of Gaseous Conduction and Electronics" (McGraw-Hill, 1941).
(7) Buchtiger, P.: "Die Dynamischen Charaktaristiken einer Bogentladung," *Helvetia Physica Acta*, 1920, **3**, p. 335.
(8) Magnold, H.: *Elektrotechnische Nachrichten*, 1940, **17**, p. 57.
(9) Druvestyn and Penning: "Electrical Discharges in Gases," *Review of Modern Physics*, 1939, **12**, p. 87.
(10) Ornstein, L. S., and Brinkman, H.: "Der Thermische Mechanismus in der Säule des Lichtbogens," *Physica*, 1934, **1**, p. 797.
(11) Ellenbass, W.: "Der Gradient der Uberhochdruck-Quecksilber Entladung," *ibid.*, 1937, **4**, p. 279.
(12) Suits, C. G.: *Physics*, 1930, **6**, pp. 190 and 315.
(13) Suits, C. G.: "Current Densities, Lumen Efficiency and Brightness in A, N_2, He and H_2 Arcs," *Journal of Applied Physics*, 1939, **10**, p. 730.
(14) Von Engel, A., and Stienbeck, M.: "Elektriche Gastentladungen."
(15) Kneser, H. O.: "The Interpretation of the Anomalous Sound-Absorption in Air and Oxygen in terms of Molecular Collisions," *Journal of the Acoustical Society of America*, 1933–34, **5**, p. 122.
(16) Compton, K. T.: *Review of Modern Physics*, 1930, **12**, p. 211.

PROCEEDINGS OF THE INSTITUTION OF ELECTRICAL ENGINEERS

ISSUED IN THREE PARTS AS FOLLOWS:
Part A. POWER ENGINEERING (*February, April, etc.*)
Part B. RADIO AND ELECTRONIC ENGINEERING (*January, March, etc.*)
Part C. INSTITUTION MONOGRAPHS (*March and September only*)

Part B. RADIO AND ELECTRONIC ENGINEERING (INCLUDING COMMUNICATION ENGINEERING)

JANUARY 1955

CONTENTS OF THIS ISSUE

PAGE
The President's Inaugural Address . . . . . J. ECCLES, C.B.E., B.Sc. 1
Radio Section: Chairman's Address . . . . . C. W. OATLEY, M.A., M.Sc. 7
The Properties of Artificial Dielectrics at Centimetre Wavelengths . . . . . J. BROWN, M.A., and WILLIS JACKSON, D.Sc., D.Phil., F.R.S. 11
Some Experiments on Artificial Dielectrics at Centimetre Wavelengths . . . . . M. M. Z. EL-KHARADLY, B.Sc., Ph.D. 17
A General Experimental Method to determine the Properties of Artificial Media at Centimetre Wavelengths, applied to an Array of Parallel Metallic Plates . . . . . R. I. PRIMICH, Ph.D., B.Sc. 26
The Relative Permittivity of Tetragonal Arrays of perfectly Conducting Thin Discs . . J. BROWN, M.A., and WILLIS JACKSON, D.Sc., D.Phil., F.R.S. 37
Discussion on "The Effect of Irradiation on the Calibration of 2 cm-Diameter Sphere-Gaps" 42
ome Features of V.H.F. Tropospheric Propagation . . . . . M. W. GOUGH, M.A. 43
oise Generation in Crystals and in Ceramic Forms of Barium Titanate when subjected to Electric Stress . . . . A. C. KIBBLEWHITE, M.Sc., Ph.D. 59
scussion on "Special Effects for Television Studio Productions" 68
ign of a Logarithmic Receiver . . . . . S. ROZENSTEIN 69
ssion on "A Method of Designing Transistor Trigger Circuits" 74
ation in Curved and Twisted Waveguides of Rectangular Cross-Section . . . . . L. LEWIN 75
n Cut-Off Magnetrons . . . . . R. C. GLASS, B.A., B.Sc., G. D. SIMS, M.Sc., B.Sc., and A. G. STAINSBY, B.Sc. 81
f Long-Distance Radio-Wave Propagation above 30 Mc/s.
W. J. BRAY, M.Sc.(Eng.), H. G. HOPKINS, B.Sc., Ph.D., F. A. KITCHEN, B.Sc., and J. A. SAXTON, D.Sc., Ph.D. 87
de-Band Dummy Load for Measuring Power at Very-High and Ultra-High Frequencies . . . . . W. HERSCH, B.Sc.(ENG.) 96
l Equipment and Techniques for a Study of Millimetre-Wave Propagation.
W. E. WILLSHAW, M.B.E., M.Sc.Tech., H. R. L. LAMONT, M.A., Ph.D., and E. M. HICKIN 99
"A Moving-Coil Relay applied to Modern High-Speed Protective Systems" 112
tution Monographs:
ower Convertors: An Analysis of Performance and Design . . . . . R. H. EVANS, B.Sc. 113
pedance of Plasma Discharges in Mercury Vapour . . . . . S. E. YUSSUF, Ph.D., and PROF. J. C. PRESCOTT, D.Eng. 115

ving.—Within the terms of the Royal Society's Declaration on Fair Copying, to which The Institution subscribes, material may be copied from (prior to 1949, the *Journal*) *which are out of print and from which reprints are not available*. The terms of the Declaration and particulars of a Photoprint Service afforded by the Science Museum Library, London, are published in the *Journal* from time to time.

-It is requested that bibliographical reference to an Institution paper should always include the serial number of the paper and the month and vill be found at the top right-hand corner of the first page of the paper. This information should precede the reference to the Volume and Part.

.—SMITH, J.: "Reflections from the Ionosphere," *Proceedings I.E.E.*, Paper No. 3001 R, December, 1954 (102 B, p. 1234).

HE BENEVOLENT FUND

applications for assistance from the Fund has shown a marked increase years, and in 1954 these fresh demands considerably exceeded the increase The state of the Fund has enabled the Court of Governors to maintain standard of assistance in the necessitous cases but they are anxious lp should not be impaired.

ed by about 40% of the members, and the Governors' best thanks ho subscribe. They do, however, specially appeal to those who te to the Fund.

tions and Donations may be sent by post to
RPORATED BENEVOLENT FUND OF
TION OF ELECTRICAL ENGINEERS,
Y PLACE, LONDON, W.C.2
ne of the Local Hon. Treasurers of the Fund

Y SUBSCRIPTIONS, DONATIONS, LEGACIES

EASURERS OF THE FUND:

ds	NORTHERN IRELAND CENTRE	*G. H. Moir, J.P.*
E.	SCOTTISH CENTRE	*R. H. Dean, B.Sc.Tech.*
	NORTH SCOTLAND SUB-CENTRE	*P. Philip*
	SOUTH MIDLAND CENTRE	*W. E. Clark*
	RUGBY SUB-CENTRE	*H. Orchard*
	SOUTHERN CENTRE	*G. D. Arden*
	WESTERN CENTRE (BRISTOL)	*A. H. McQueen*
	STERN CENTRE (CARDIFF)	*D. J. Thomas*
	TH-WESTERN SUB-CENTRE	*W. E. Johnson*

le Bar 7676. *Telegrams:* "Voltampere, Phone, London."
king and London.